

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 15:17:16 ; Search time 2995.75 Seconds
(without alignments)
170.772 Million cell updates/sec

Title: US-10-725-373-2

Perfect score: 45

Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10725373/runat_01122005_114445_21447/app.query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373 @CGN_1_1_5655 @runat_01122005_114445_21447 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*

1: gb.ba.*

2: gb.in.*

3: gb.env.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: gb.htg.*

15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	27	6	BD131676 Carcinoem
2	45	100.0	27	6	CS089178 Sequence
3	45	100.0	27	6	AR560605 Sequence

4	45	100.0	2106	6	AX133657	Sequence
5	45	100.0	2106	6	AX192349	Sequence
6	45	100.0	2106	6	AX393888	Sequence
7	43	95.6	27	6	BD131677	Carcinoem
8	43	95.6	27	6	CS089179	Sequence
9	43	95.6	27	6	AR560606	Sequence
c 10	42	93.3	196838	14	AC115196	AC115196 Rattus no
c 11	42	93.3	282895	14	AC096056	AC096056 Rattus no
c 12	41	91.1	110000	1	CP000099_34	Continuation (35 o
c 13	41	91.1	110000	1	CP000099_35	Continuation (36 o
14	40	88.9	27	6	BD131675	BD131675 Carcinoem
15	40	88.9	27	6	CS089177	Sequence
16	40	88.9	27	6	AR560604	Sequence
c 17	40	88.9	80	6	AB7744	AB7744 Sequence 38
c 18	40	88.9	80	6	BD195831	BD195831 Method fo
19	40	88.9	155	6	AX193221	Sequence
20	40	88.9	256	6	AX260775	Sequence
c 21	40	88.9	407	6	AX260391	Sequence
c 22	40	88.9	409	6	AR273297	Sequence
c 23	40	88.9	409	6	AR273585	Sequence
c 24	40	88.9	409	6	AR273719	Sequence
c 25	40	88.9	409	6	AR276878	Sequence
c 26	40	88.9	409	6	AR277166	Sequence
c 27	40	88.9	409	6	AR277300	Sequence
c 28	40	88.9	409	6	AR407153	Sequence
c 29	40	88.9	409	6	AR407441	Sequence
c 30	40	88.9	409	6	AR407575	Sequence
c 31	40	88.9	409	6	AR441003	Sequence
c 32	40	88.9	409	6	AR441291	Sequence
c 33	40	88.9	409	6	AR441425	Sequence
c 34	40	88.9	409	6	AR543814	Sequence
c 35	40	88.9	409	6	AR544102	Sequence
c 36	40	88.9	409	6	AR544236	Sequence
c 37	40	88.9	409	6	AR638972	Sequence
c 38	40	88.9	409	6	AR639260	Sequence
c 39	40	88.9	409	6	AR639394	Sequence
c 40	40	88.9	409	6	AX260450	Sequence
c 41	40	88.9	409	6	AX351472	Sequence
c 42	40	88.9	409	6	AX368330	Sequence
c 43	40	88.9	409	6	AX368330	Sequence
c 44	40	88.9	409	6	AX368618	Sequence
c 45	40	88.9	409	6	AX368752	Sequence
					AX396625	Sequence

ALIGNMENTS

RESULT 1	BD131676	BD131676	27 bp	DNA	linear	PAT 18-SEP-2002
LOCUS	BD131676	Carcinoembryonic antigen (CEA)	agonist and antagonist peptides.			
DEFINITION	BD131676	BD131676.1	GI:23226621			
ACCESSION	BD131676	JP 2002500002-A/2.				
VERSION	BD131676.1	GI:23226621				
KEYWORDS	JP 2002500002-A/2.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;					
TITLE	Hominidae; Homo.					
JOURNAL	1 (bases 1 to 27)					
COMMENT	Schlom, J., Barzaga, E. and Zaremba, S.					
	Carcinoembryonic antigen (CEA) agonist and antagonist peptides					
	Patent: JP 2002500002-A 2 08-JAN-2002;					
	THE UNITED STATES OF AMERICA					
	OS Homo sapiens (human)					
	PN JP 2002500002-A/2					
	PD 08-JAN-2002					
	PF 22-SEP-1998					
	PR 10-OCT-1997					
	PI JEFFREY SCHLOM, ELENE BARZAGA, SAM ZAREMBA					
	PC C12N15/09, A61K38/00, A61K45/00, A61K48/00, A61P35/00, A61P37/02,					
	PC A61P43/00,					
	PC C07K14/705, C07K16/28, C12N5/10, C12N15/00, A61K37/02, C12N5/00 CC					
	Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH					

Key Location/Qualifiers
FT source 1..27
FT /organism="Homo sapiens (human)"

FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.0139 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x BD131676 (1-27)

Qy 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
Db 1 TACCTTCGGGAGCGGACCTCAACCTC 27

RESULT 2
CS089178 27 bp DNA linear PAT 25-MAY-2005
LOCUS
DEFINITION Sequence 7 from Patent EP1447414.
ACCESSION CS089178
VERSION CS089178.1 GI:66714457
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
1
AUTHORS Schlom, J., Salazar, M.E. and Zaremba, S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: EP 1447414-A 7 18-AUG-2004;
Department of Health and Human Services (US)

FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.0139 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x CS089178 (1-27)

Qy 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
Db 1 TACCTTCGGGAGCGGACCTCAACCTC 27

RESULT 3
LOCUS
DEFINITION Sequence 7 from patent US 6756038.
ACCESSION AR560605
VERSION AR560605.1 GI:53972926
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 27)

AUTHORS Schlom, J., Barzaga, E. and Zaremba, S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 7 29-JUN-2004;
The United States of America as represented by the Department of Health and Human Services; Washington, DC;
WOX;

FEATURES
source 1..27
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0.0139 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AR560605 (1-27)

Qy 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
Db 1 TACCTTCGGGAGCGGACCTCAACCTC 27

RESULT 4
AX133657 2106 bp DNA linear PAT 15-MAY-2001
LOCUS
DEFINITION Sequence 111 from Patent WO0130382.
ACCESSION AX133657
VERSION AX133657.1 GI:14139699
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE
1
AUTHORS Berinstein, N., Tartaglia, J., Moingeon, P. and Barber, B.
TITLE Method of inducing and/or enhancing an immune response to tumor antigens
JOURNAL Patent: WO 0130382-A 111 03-MAY-2001;
Aventis Pasteur Limited (CA)

FEATURES
source 1..2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"

CDS
1..2106
/notes="unnamed protein product; modified CEA"

/codon_start=1
/transl_table=11
/protein_id="CAC38955.1"
/db_xref="GI:14139700"
/translations="MESPAPPHRWCTPQRLLLTASLLTFWNPPTTAKLTISTPFPN
VAGKEVLLVHNLPHQLFGYSWKGVNDRQIIGVIGTQATPGPAYSGREIIV
PNASLLIONDTGFTLHVKSLDNEATGQPRVYPELKPSSISNNKPEVDK
DAVAFCEPETQATYLMWVNSQLPVPSPRLQSLNGRNTLTENVTRNDTASYKCEIQ
NPVSARRSDVILNVLGPDPTISPLTYSRGENLNLSCHAASNPPAQYSWFVNGT
FQOSTQELFIPNITVNNSGSYTCQAHNSDTGLNRTVTITVYEPKPFITSNNSPV
EDEDVALTCEPEIQNTYLMWVNSQLPVPSPRLQSLNGRNTLTLSVTRNDVGPVEK
GIGNELSDVHSDPVIINVLGPDPTISPSYTYRPGVNLNLSCHAAASNPPAQYSWLI
DGNIQHTBELFISNTEKNSGLYTCQANNSAGHSRTTYKTIIVSAELKPSSISNN
SKPEVDKDAVAFCEPEAQNTYLMWVNSQLPVPSPRLQSLNGRNTLTENVTRNDAR
AYVCGTQNSVANRSDPVLIDVLPDPTIISPDSSYLSGADNLNLSCHASASNPPQY
SWRINGIPQHTQVLFIAKTIPTNNNGTYACFVSNLATGRNNSIVKTSITVASGTSPLG
SAGATVGIMIGVLGVALLI"

ORIGIN
Alignment Scores:
Pred. No.: 1.86 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AX133657 (1-2106)
QY 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 5
AX192349
LOCUS AX192349 2106 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 3 from Patent WO0149317.
ACCESSION AX192349
VERSION AX192349.1 GI:15210326
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Entage,P., Barber,B.H., Sambhara,S. and Sia,C.D.
TITLE Enhancing the immune response to an antigen by presensitizing with
an inducing agent prior to immunizing with the inducing agent and
the antigen
JOURNAL Patent: WO 0149317-A 3 12-JUL-2001;
Aventis Pasteur Limited (CA)
FEATURES
source
1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1. .2106
/codon_start=1
/translate=1
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1. .2106
/feature="unnamed protein product; modified CEA"

CDS
1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1. .2106
/feature="unnamed protein product; modified CEA"

US-10-725-373-2 (1-9) x AX192349 (1-2106)
QY 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 6
AX393888
LOCUS AX393888 2106 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 2 from Patent WO0210379.
ACCESSION AX393888
VERSION AX393888.1 GI:19701852
KEYWORDS
```

```

SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berinstein,N., Tartaglia,J., Tine,J.A., Panicali,D.L., Gritz,L. and
Schlom,J.
TITLE Modified cea and uses thereof
JOURNAL Patent: WO 0210379-A 2 07-FEB-2002;
Aventis Pasteur Limited (CA) ; Therion Biologics (US) ; National
Cancer Institute (US)
FEATURES
source
1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/feature="Cea modified polypeptide"

ORIGIN
Alignment Scores: 1.86 Length: 2106
Pred. No.: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AX393888 (1-2106)
QY 1 TyrLeuSerGlyAlaAspLeuAenLeu 9
DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 7
BD131677
LOCUS BD131677 27 bp DNA linear PAT 18-SEP-2002
DEFINITION Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION BD131677
VERSION BD131677.1 GI:23226622
KEYWORDS JP 2002500002-A/3.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Baraga,E. and Zaremba,S.
TITLE Carcinoembryonic antigen (CEA) agonist and antagonist peptides
JOURNAL Patent: JP 2002500002-A 3 08-JAN-2002;
THE UNITED STATES OF AMERICA
COMMENT OS Homo sapiens (human)
PN JP 2002500002-A/3
PD 08-JAN-2002
PF 22-SEP-1998 JP 2000516030
PI 10-OCT-1997 US 60/061589
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
A61P43/00,
PC A61P43/00,
PC C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
Key Location/Qualifiers
FT source 1. .27
FT /organism="Homo sapiens (human)".

FEATURES
source
1. .27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores: 0.0432 Length: 27
Pred. No.: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
```


Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 196838)
 Worley,K.C.
 Direct Submission
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 196838)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23618799.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GSYI
 Center clone name: CH230-291E8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 175442 bases at least Q40
 Consensus quality: 177660 bases at least Q30
 Consensus quality: 179434 bases at least Q20
 Estimated insert size: 177868; sum-of-contigs estimation
 Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 39826: contig of 39826 bp in length
 * 39827 39926: gap of unknown length
 * 39927 194015: contig of 154089 bp in length
 * 194016 194115: gap of unknown length
 * 194116 195420: contig of 1305 bp in length
 * 195421 195520: gap of unknown length
 * 195521 196838: contig of 1318 bp in length.

Location/Qualifiers
 1. .196838
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-291E8"
 /notes="complement(32622..33371)
 /notes="clone boundary
 clone_end:Sp6
 site:
 end_sequence:B2166109"

misc_feature
 source

misc_feature 35496..36601
 /note="wgs contig"
 38810..39826
 /note="wgs contig"
 39827..39926
 /estimated_length=unknown
 39927..43082
 /note="wgs contig"
 194016..194115
 /estimated_length=unknown
 195421..195520
 /estimated_length=unknown
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.68e+03 Length: 196838
 Score: 42.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 93.33% Indels: 0
 DB: 14 Gaps: 0
 US-10-725-373-2 (1-9) x AC115196 (1-196838)
 Qy 1 TyrluSerGlyAlaAspLeuAsnLeu 9
 |||||:|||||||
 Db 162369 TACCTCTCAGGTTTCAGACTTAACCTA 162343
 RESULT 11
 AC096056/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-22P2, WORKING DRAFT SEQUENCE, 2
 unorderd pieces.
 AC096056 GI:24817935
 VERSION AC096056.5 HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.
 1 (bases 1 to 282895)
 Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J.,
 Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D.,
 AyalaBebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flaggs,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,K., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensuhewa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

CP000099_27 2700001 2810000
CP000099_28 2800001 2910000
CP000099_29 2900001 3010000
CP000099_30 3000001 3110000
CP000099_31 3100001 3210000
CP000099_32 3200001 3310000
CP000099_33 3300001 3410000
CP000099_34 3400001 3510000
CP000099_35 3500001 3610000
CP000099_36 3600001 3710000
CP000099_37 3700001 3810000
CP000099_38 3800001 3910000
CP000099_39 3900001 4010000
CP000099_40 4000001 4110000
CP000099_41 4100001 4210000
CP000099_42 4200001 4310000
CP000099_43 4300001 4410000
CP000099_44 4400001 4510000
CP000099_45 4500001 4610000
CP000099_46 4600001 4710000
CP000099_47 4700001 4810000
CP000099_48 4800001 4837408

Continuation (35 of 49) of CP000099 from base 3400001 (CP000099 Methanosarcina barkeri

Alignment Scores:
Pred. No.: 1.54e+03 Length: 110000
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 91.11% Indels: 0
DB: 1 Gaps: 0

US-10-725-373-2 (1-9) x CP000099_34 (1-110000)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
DB 109137 TTTTATCAGGACGAGATTAAATTAA 109163

RESULT 13
CP000099_35
WPCOMMENT
Sequence split into 49 fragments LOCUS CP000099 Accession CP000099

Fragment Name	Begin	End
CP000099_00	1	110000
CP000099_01	100001	210000
CP000099_02	200001	310000
CP000099_03	300001	410000
CP000099_04	400001	510000
CP000099_05	500001	610000
CP000099_06	600001	710000
CP000099_07	700001	810000
CP000099_08	800001	910000
CP000099_09	900001	1010000
CP000099_10	1000001	1110000
CP000099_11	1100001	1210000
CP000099_12	1200001	1310000
CP000099_13	1300001	1410000
CP000099_14	1400001	1510000
CP000099_15	1500001	1610000
CP000099_16	1600001	1710000
CP000099_17	1700001	1810000
CP000099_18	1800001	1910000
CP000099_19	1900001	2010000
CP000099_20	2000001	2110000
CP000099_21	2100001	2210000
CP000099_22	2200001	2310000
CP000099_23	2300001	2410000
CP000099_24	2400001	2510000
CP000099_25	2500001	2610000
CP000099_26	2600001	2710000
CP000099_27	2700001	2810000
CP000099_28	2800001	2910000
CP000099_29	2900001	3010000

CP000099_30 3000001 3110000
CP000099_31 3100001 3210000
CP000099_32 3200001 3310000
CP000099_33 3300001 3410000
CP000099_34 3400001 3510000
CP000099_35 3500001 3610000
CP000099_36 3600001 3710000
CP000099_37 3700001 3810000
CP000099_38 3800001 3910000
CP000099_39 3900001 4010000
CP000099_40 4000001 4110000
CP000099_41 4100001 4210000
CP000099_42 4200001 4310000
CP000099_43 4300001 4410000
CP000099_44 4400001 4510000
CP000099_45 4500001 4610000
CP000099_46 4600001 4710000
CP000099_47 4700001 4810000
CP000099_48 4800001 4837408

Continuation (36 of 49) of CP000099 from base 3500001 (CP000099 Methanosarcina barkeri

Alignment Scores:
Pred. No.: 1.54e+03 Length: 110000
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 91.11% Indels: 0
DB: 1 Gaps: 0

US-10-725-373-2 (1-9) x CP000099_35 (1-110000)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
DB 9137 TTTTATCAGGACGAGATTAAATTAA 9163

RESULT 14

BD131675
LOCUS BD131675 27 bp DNA linear PAT 18-SEP-2002
DEFINITION Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION BD131675
VERSION BD131675.1 GI:23226620
KEYWORDS JP 2002500002-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlow,J., Barzaga,E. and Zaremba,S.
TITLE Carcinoembryonic antigen (CEA) agonist and antagonist peptides
JOURNAL Patent: JP 2002500002-A 1 08-JAN-2002;
THE UNITED STATES OF AMERICA

COMMENT
OS Homo sapiens (human)
PN JP 2002500002-A/1
PD 08-JAN-2002
PF 22-SEP-1998 JP 2000516030
PR 10-OCT-1997 US 60/061589
PI JEFFREY SCHLOW, ELENE BARZAGA, SAM ZAREMBA
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
PC A61P43/00,
PC C07K14/705,C07K16/28,C12N5/10,C12N15/00,A61K37/02,C12N5/00 CC
Carcinoembryonic antigen (CEA) agonist and antagonist peptides PH
Key Location/Qualifiers
FT source 1..27
FT /organism='Homo sapiens (human)'.
Location/Qualifiers
1..27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source

ORIGIN

Alignment Scores:

Pred. No.: 0.238 Length: 27
Score: 40.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 88.89% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x BD131675 (1-27)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
|||||
DB 1 TACCTTCGGGAGCGAACCTCAACCTC 27

RESULT 15

CS089177 LOCUS 27 bp DNA linear PAT 25-MAY-2005
DEFINITION Sequence 6 from Patent EP1447414.
ACCESSION CS089177
VERSION CS089177.1 GI:66714456
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE

1 Schlom, J., Salazar, M.E. and Zarembo, S.
AUTHORS Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
TITLE Patent: EP 1447414-A 6 18-AUG-2004;
JOURNAL Department of Health and Human Services (US)

FEATURES

source
1..27
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 0.238 Length: 27
Score: 40.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 88.89% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x CS089177 (1-27)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
|||||
DB 1 TACCTTCGGGAGCGAACCTCAACCTC 27

Search completed: December 6, 2005, 19:50:35
Job time : 3059.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 10:17:58 ; Search time 367.25 Seconds
(without alignments)
163.328 Million cell updates/sec

Title: US-10-725-373-2
Perfect score: 45
Sequence: 1 YLSGADLNL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10725373/runat_01122005_114444_21420/app.query.fasta_1.796
-DB=N Geneseq -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373 @CGN_1_11244 @runat_01122005_114444_21420 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq_21:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	27	2 AAX56258	Aax56258 Carcinoem
2	45	100.0	2105	6 Aai72497	Aai72497 CEA agoni
3	45	100.0	2106	4 Aah20121	Aah20121 Modified
4	45	100.0	2106	5 Aad07347	Aad07347 Modified

5	45	100.0	2106	6	AAI72489	AAI72489 CEA agoni
6	45	100.0	2106	10	ADEI3860	ADEI3860 CEA-CAR6D
7	45	100.0	2106	10	ADEI3861	ADEI3861 CEA (6D)-1
8	45	100.0	2106	14	ADZ58977	ADZ58977 Novel CEA
9	45	100.0	2106	14	ADZ58978	ADZ58978 Novel CEA
10	45	100.0	2106	14	AEAI1047	AEAI1047 DNA encod
11	45	100.0	2106	14	AEAI1047	AEAI1047 Wobbled C
12	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
13	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
14	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
15	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
16	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
17	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
18	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
19	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
20	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
21	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
22	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
23	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
24	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
25	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
26	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
27	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
28	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
29	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
30	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
31	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
32	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
33	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
34	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
35	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
36	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
37	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
38	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
39	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
40	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
41	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
42	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
43	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
44	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant
45	45	100.0	2106	14	AEAI2740	AEAI2740 Multi-ant

ALIGNMENTS

RESULT 1
AAX56258
ID AAX56258 standard; DNA; 27 BP.
XX
AC AAX56258;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:7.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
FN MO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX

PI Schlom J, Barzaga E, Zaremba S;
 XX WPI; 1999-326544/27.
 XX Peptide agonists and antagonists of carcinoembryonal antigen.
 XX Claim 22; Page 20; 72pp; English.
 XX The present invention describes peptides (A) that comprise agonists (Ia) or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are used in vaccines to kill or inhibit carcinoma cells that express CEA or its epitopes, particularly for treating gastrointestinal, breast, pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also be used to proliferate T cells, e.g. from vaccinated subjects, for use in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction to cancer immunotherapy (i.e. to prevent attack on normal but CEA-expressing cells). (Ia) are more active than native sequence (I) and generate a highly specific and systemic anti-CEA response. Cytotoxic T cells generated recognize both (Ia) and native CEA epitopes. The present sequence encodes a specifically claimed example of (Ia)
 XX Sequence 27 BP; 5 A; 10 C; 6 G; 6 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 0.059 Length: 27
 Score: 45.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-725-373-2 (1-9) x AAX56258 (1-27)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
 Db 1 TACCTTTCGGAGCGACCTCAACCTC 27

RESULT 2
 AAI72497
 ID AAI72497 standard; DNA; 2105 BP.
 XX AAI72497;
 XX 16-MAY-2002 (first entry)
 DE CEA agonist coding sequence #2.
 XX CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response; carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung; prostate; cancer; gene; ss.
 XX Synthetic.
 XX Key Location/Qualifiers
 FH CDS 1..2105
 FT /*tag= a
 FT /product= "CEA agonist polypeptide"
 XX WO200210379-A2.
 XX 07-FEB-2002.
 XX 27-JUL-2001; 2001WO-CA001092.
 XX 31-JUL-2000; 2000US-0222043P.
 XX (AVET) AVENTIS PASTEUR LTD.
 PA (THER-) THERION BIOLOGICS.
 PA (USSH) US NAT CANCER INST.
 XX Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
 PI Schlom J;

XX WPI; 2002-206189/26.
 XX Carcinoembryonic antigen agonist polypeptide for inducing an immune response in animal against antigen and for inhibiting an epitope antigen expressing carcinoma cell, comprises a modified antigen epitope.
 XX Claim 4; Page 66-67; 69pp; English.
 XX This sequence encodes the carcinoembryonic antigen (CEA) agonist polypeptide of the invention. This sequence represents the sequence given in the Seq ID listing in the specification, and does not directly encodes the CEA agonist polypeptide given in AAB47919. The CEA agonist contains a modified CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to increase its immunogenicity. The CEA agonist polypeptide of the invention, or DNA encoding it, are useful for: (i) inducing an immune response in an animal directed against a CEA protein or fragment, CEA agonist, a CEA epitope, a modified CEA epitope, cells expressing or binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope expressing carcinoma cell, which is a gastrointestinal, breast, pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a patient, hence is useful for manufacture of a medicament for the treatment of cancer
 XX Sequence 2105 BP; 555 A; 658 C; 441 G; 451 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 8.03 Length: 2105
 Score: 45.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AAI72497 (1-2105)

QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
 Db 1809 TACCTTTCGGAGCGACCTCAACCTC 1835

RESULT 3
 AAH20121
 ID AAH20121 standard; cDNA; 2106 BP.
 XX AAH20121;
 XX 08-AUG-2001 (first entry)
 DE Modified Carcinoembryonic antigen (CEA) encoding cDNA SEQ ID NO:111.
 DE Virus; adenovirus; poxvirus; alphavirus; immune response; gp100; tumour antigen; CEA; carcinoembryonic antigen; immunostimulant; cytostatic; immunotherapy; interferon-gamma; IFN-gamma; cancer; ss.
 XX Unidentified.
 OS WO200130382-A1.
 XX 03-MAY-2001.
 XX 20-OCT-2000; 2000WO-CA001253.
 XX 22-OCT-1999; 99US-0160879P.
 PR 07-AUG-2000; 2000US-0223325P.
 XX (AVET) AVENTIS PASTEUR LTD.
 XX Berinstein N, Tartaglia J, Moingeon P, Barber B;
 WPI; 2001-308587/32.
 DR P-PSDB; AAB97817.
 XX Inducing immune response to tumor antigen, useful in immunotherapy of

PT cancer, by administering the antigen to a lymphatic site.

PS Disclosure; Fig 8; 60pp; English.

XX
CC The present invention describes a method for inducing an immune response, in an animal, to a tumour antigen (Ag) comprising administering Ag, or nucleic acid (I) that encodes it, to a lymphatic site. Cynomolgus monkeys (Macaca fascicularis) were injected with a modified form of gp100 antigen (a) into the left inguinal lymph node or (b) subcutaneously. Both animals of (a) developed a cell-mediated response (indicated by production of interferon-gamma from T lymphocytes when exposed to gp100 peptides), but only 2 of 4 animals of (b) did so. Also animals in (a) produced a far greater antibody response to gp100. The method is used in immunotherapy of a wide range of cancers through induction of a specific immune response (humoral and cellular) against the tumour antigens. When administered to a lymphatic site, Ag (or (I)) induces a stronger immune response than administration by other routes and may also break tolerance to Ag. AAB97708 and AAB97709 represent gp100 epitopes; AAB97710 to AAB97815 represent peptides derived from gp100 which stimulate interferon (IFN)-gamma production; AAB97816 to AAB97817 encodes the modified gp100 protein given in AAB97816; AAB97818 encodes the modified carcinoembryonic antigen (CEA) protein given in AAB97817; and AAB97818 represents a CEA modified antigen peptide, all of which are used in the exemplification of the present invention

XX
SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-725-373-2 (1-9) x AAB97818 (1-2106)

Qy 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 4

AAD07347

ID AAD07347 standard; DNA; 2106 BP.

XX AAD07347;

XX 18-SEP-2001 (first entry)

DE Modified carcinoembryonic antigen (CEA) DNA.

XX Tumour-associated antigen; TAA; cytostatic; vaccine; gene therapy;
KW immune response; tetanus toxoid; TT; diphtheria toxoid; DT; prophylactic;
KW cancer; therapeutic; carcinoembryonic antigen; CEA; ds.

XX Synthetic.

XX Key Location/Qualifiers
FH 1. .2106
CDS /*tag= a

FT /product= "Modified carcinoembryonic antigen (CEA) "

XX WO200149317-A2.

PN 12-JUL-2001.

XX 05-JAN-2001; 2001WO-CA000005.

XX 05-JAN-2000; 2000US-0174587P.

XX (AVET) AVENTIS PASTEUR LTD.

XX Entage P, Barber BH, Sambhara S, Sia CDY;

XX WPI; 2001-441790/47.
DR P-PSDB; AAB05117.

XX
PT Enhancing immune response to antigen such as tumor antigen for treating cancer in an animal involves administering an inducing agent to the animal followed by administering inducing agent-antigen mixture.

XX Claim 9; Fig 3; 62pp; English.

XX The invention relates to a method of enhancing an immune response against tumour-associated antigens (TAAs), such as gp100 and carcinoembryonic antigen (CEA) in an animal. The method involves priming of the animal with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid (DT), subsequently followed by administration of an inducing agent-antigen mixture. The method provides the enhancement or augmentation of the immune response to the antigen and/or improves a vaccination protocol by allowing use of less antigen. The immunisation of the animal with tumour-associated antigen is useful for the prophylactic or therapeutic treatment of cancer. The present DNA sequence encodes modified carcinoembryonic antigen (CEA) related to the invention

XX Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-725-373-2 (1-9) x AAD07347 (1-2106)

Qy 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9

Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 5

AAT72489

ID AAT72489 standard; DNA; 2106 BP.

XX AAT72489;

XX 16-MAY-2002 (first entry)

DE CEA agonist coding sequence #1.

XX CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
KW prostate; cancer; gene; ds.

XX Synthetic.

XX Key Location/Qualifiers
FH 1. .2106
CDS /*tag= a

FT /product= "CEA agonist polypeptide"

XX WO200210379-A2.

PN 07-FEB-2002.

XX 27-JUL-2001; 2001WO-CA001092.

XX 31-JUL-2000; 2000US-0222043P.

XX (AVET) AVENTIS PASTEUR LTD.

XX (THER-) THERION BIOLOGICS.

XX (USSH) US NAT CANCER INST.

XX Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;

PI Schlom J;

XX WPI; 2002-206189/26.
 DR P-PSDB; AAB47918.
 XX Carcinoembryonic antigen agonist polypeptide for inducing an immune
 PT response in animal against antigen and for inhibiting an epitope antigen
 PT expressing carcinoma cell, comprises a modified antigen epitope.
 XX Claim 4; Fig 1; 69pp; English.
 XX This sequence encodes the carcinoembryonic antigen (CEA) agonist
 CC polypeptide of the invention. This sequence represents the sequence given
 CC in the figures in the specification, and it directly encodes the CEA
 CC agonist polypeptide given in AAB47918. The CEA agonist contains a
 CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp
 CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the
 CC invention, or DNA encoding it, are useful for: (i) inducing an immune
 CC response in an animal directed against a CEA protein or fragment, CEA
 CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
 CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
 CC expressing carcinoma cell, which is a gastrointestinal, breast,
 CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
 CC patient, hence is useful for manufacture of a medicament for the
 CC treatment of cancer
 XX
 SQ Sequence 2106 BP; 559 A; 658 C; 442 G; 447 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.03 Length: 2106
 Score: 45.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AAI72489 (1-2106)
 QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
 DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 6
 ADEI3860
 ID ADEI3860 standard; DNA; 2106 BP.
 XX
 AC ADEI3860;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE CEA-CAP6D nucleotide sequence SEQ ID NO:23.
 XX
 KW carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
 KW tumour antigen; immunotherapy; gene; ds.
 XX
 OS Unidentified.
 XX
 PN WO2003085087-A2.
 XX
 PD 16-OCT-2003.
 XX
 PF 09-APR-2003; 2003WO-US010916.
 XX
 PR 09-APR-2002; 2002US-0372972P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 PA (THER-) THERION BIOLOGICS INC.
 XX
 PI Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
 XX
 DR WPI; 2003-877029/81.
 XX
 PT New isolated DNA molecule comprising the carcinoembryonic antigen (6D) -
 PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
 PT determining the effectiveness of a chemotherapeutic or other treatment

PT determining the effectiveness of a chemotherapeutic or other treatment
 PT regimen.
 XX
 PS Example 1; SEQ ID NO 23; 56pp; English.
 XX
 CC The present invention describes an isolated DNA molecule comprising the
 CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
 CC ADEI3861), or its fragment. Also described: (1) an expression vector
 CC comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
 CC describes above; (2) a composition comprising the expression vector of
 CC (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
 CC comprising administering to a host the expression vector of (1). CEA(6D)-
 CC 1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
 CC nucleic acid and target polypeptide are useful for diagnosing, preventing
 CC and treating cancer, predicting prognosis, or determining the
 CC effectiveness of a chemotherapeutic or other treatment regimen. The
 CC expression vector may be used for the insertion and expression of CEA(6D)
 CC -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
 CC treatment of cancer. The target polypeptides are useful in generating
 CC antibodies used in screening assays or for immunotherapy. The present
 CC sequence represents the CEA-CAP6D nucleotide sequence, which is given in
 CC comparison with CEA(6D)-1,2 in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.03 Length: 2106
 Score: 45.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-2 (1-9) x ADEI3860 (1-2106)
 QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
 DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 7
 ADEI3861
 ID ADEI3861 standard; DNA; 2106 BP.
 XX
 AC ADEI3861;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE CEA(6D)-1,2 nucleotide sequence SEQ ID NO:24.
 XX
 KW carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
 KW tumour antigen; immunotherapy; gene; ds.
 XX
 OS Unidentified.
 XX
 PN WO2003085087-A2.
 XX
 PD 16-OCT-2003.
 XX
 PF 09-APR-2003; 2003WO-US010916.
 XX
 PR 09-APR-2002; 2002US-0372972P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 PA (THER-) THERION BIOLOGICS INC.
 XX
 PI Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
 XX
 DR WPI; 2003-877029/81.
 XX
 PT New isolated DNA molecule comprising the carcinoembryonic antigen (6D) -
 PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
 PT determining the effectiveness of a chemotherapeutic or other treatment

PT regimen.
PS Claim 1; SEQ ID NO 24; 56pp; English.
XX
CC The present invention describes an isolated DNA molecule comprising the
CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
CC ADE13861), or its fragment. Also described: (1) an expression vector
CC comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
CC describes above; (2) a composition comprising the expression vector of
CC (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
CC comprising administering to a host the expression vector of (1). CEA(6D)-
CC 1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
CC nucleic acid and target polypeptide are useful for diagnosing, preventing
CC and treating cancer, predicting prognosis, or determining the
CC effectiveness of a chemotherapeutic or other treatment regimen. The
CC expression vector may be used for the insertion and expression of CEA(6D)
CC -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
CC treatment of cancer. The target polypeptides are useful in generating
CC antibodies used in screening assays or for immunotherapy. The present
CC sequence represents the CEA(6D)-1,2 nucleotide sequence, which is given
CC in the exemplification of the present invention.
XX
SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-725-373-2 (1-9) x ADE13861 (1-2106)

QY 1 TyLeuSerGlyAlaAspLeuAsnLeu 9
DB 1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 8
ADZ58977
ID ADZ58977 standard; DNA; 2106 BP.
XX
AC ADZ58977;
XX
DT 30-JUN-2005 (first entry)
XX
DE Novel CEA(6)-1,2 expression vector-related mCEA DNA sequence SeqID7.
XX expression; vector; CEA(6)-1,2; cytostatic; gene therapy; cancer; ds.
XX Unidentified.
XX WO2005035773-A2.
XX 21-APR-2005.
XX
XX 06-OCT-2004; 2004WO-US033145.
XX
XX 08-OCT-2003; 2003US-0509593P.
XX
XX (AVET) AVENTIS PASTEUR INC.
XX (THER-) THERION BIOLOGICS INC.
XX
XX Parrington M, Zhang L, Rovinski B, Gritz L, Greenhalgh P;
XX WPI; 2005-296285/30.
XX
XX New expression vector comprising the nucleic acid sequence CEA(6)-1,2 or
XX its fragment, useful for preventing or treating cancer.
XX
XX Example 1; SEQ ID NO 7; 72pp; English.
XX
XX This invention relates to a novel expression vector containing the
XX nucleic acid sequence CEA(6)-1,2. The invention may be useful for the
XX development of compounds with a cytostatic activity whilst the disclosed
XX sequences may be useful for gene therapy. The expression vector is useful
XX for preventing or treating cancer. The present sequence is that of an
XX mCEA expression vector DNA sequence which is related to the invention.
XX Note: Another sequence was labelled as SeqID7 in the examples of the
XX specification.
XX
SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-725-373-2 (1-9) x ADZ58977 (1-2106)

QY 1 TyLeuSerGlyAlaAspLeuAsnLeu 9
DB 1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 9
ADZ58978
ID ADZ58978 standard; DNA; 2106 BP.
XX
AC ADZ58978;
XX
DT 30-JUN-2005 (first entry)
XX
DE Novel CEA(6)-1,2 expression vector-related mCEA DNA sequence SeqID8.
XX expression; vector; CEA(6)-1,2; cytostatic; gene therapy; cancer; ds.
XX Unidentified.
XX WO2005035773-A2.
XX 21-APR-2005.
XX
XX 06-OCT-2004; 2004WO-US033145.
XX
XX 08-OCT-2003; 2003US-0509593P.
XX
XX (AVET) AVENTIS PASTEUR INC.
XX (THER-) THERION BIOLOGICS INC.
XX
XX Parrington M, Zhang L, Rovinski B, Gritz L, Greenhalgh P;
XX WPI; 2005-296285/30.
XX
XX New expression vector comprising the nucleic acid sequence CEA(6)-1,2 or
XX its fragment, useful for preventing or treating cancer.
XX
XX Example 1; SEQ ID NO 8; 72pp; English.
XX
XX This invention relates to a novel expression vector containing the
XX nucleic acid sequence CEA(6)-1,2. The invention may be useful for the
XX development of compounds with a cytostatic activity whilst the disclosed
XX sequences may be useful for gene therapy. The expression vector is useful
XX for preventing or treating cancer. The present sequence is that of an
XX mCEA expression vector DNA sequence which is related to the invention.
XX Note: Another sequence was labelled as SeqID8 in the examples of the
XX specification.
XX
SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-725-373-2 (1-9) x ADZ58977 (1-2106)

QY 1 TyLeuSerGlyAlaAspLeuAsnLeu 9
DB 1810 TACCTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 9
ADZ58978
ID ADZ58978 standard; DNA; 2106 BP.
XX
AC ADZ58978;
XX
DT 30-JUN-2005 (first entry)
XX
DE Novel CEA(6)-1,2 expression vector-related mCEA DNA sequence SeqID8.
XX expression; vector; CEA(6)-1,2; cytostatic; gene therapy; cancer; ds.
XX Unidentified.
XX WO2005035773-A2.
XX 21-APR-2005.
XX
XX 06-OCT-2004; 2004WO-US033145.
XX
XX 08-OCT-2003; 2003US-0509593P.
XX
XX (AVET) AVENTIS PASTEUR INC.
XX (THER-) THERION BIOLOGICS INC.
XX
XX Parrington M, Zhang L, Rovinski B, Gritz L, Greenhalgh P;
XX WPI; 2005-296285/30.
XX
XX New expression vector comprising the nucleic acid sequence CEA(6)-1,2 or
XX its fragment, useful for preventing or treating cancer.
XX
XX Example 1; SEQ ID NO 8; 72pp; English.
XX
XX This invention relates to a novel expression vector containing the

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-2 (1-9) x ADZ58978 (1-2106)

QY 1 TyrLeuSerGlyAlaAspLeuAnLeu 9

Db 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 10

AEA11047

ID AEA11047 standard; DNA; 2106 BP.

XX AC AEA11047;

XX DT 28-JUL-2005 (first entry)

XX DE DNA encoding wobbled-CEA (wCEA).

XX KW pancreas tumor; cytostatic; gastrointestinal disease; neoplasm; cancer;

XX KW cytostatic; immune stimulation; gene therapy; immunotherapy; ds; gene;

XX KW CEA; wCEA; carcinoembryonic antigen.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key

XX CDS

XX Location/Qualifiers

XX 988..2106

XX /*tag= a

XX /product= "wCEA"

XX /partial

XX /note= "No start codon"

XX /transl_except= (pos:1825..1827,aa:Asn)

XX PN WO2005046622-A2.

XX PD 26-MAY-2005.

XX PF 12-NOV-2004; 2004WO-US038643.

XX PR 12-NOV-2003; 2003US-0519354P.

XX PA (THER-) THERION BIOLOGICS CORP.

XX PI Panicali DL, Mazzara GP, Gritz LR;

XX DR WPI; 2005-366921/37.

XX P-PSDB; AEA11048.

XX Inducing an immunological response against a malignant pancreatic cell in

XX an individual, useful for treating pancreatic cancer, by administering

XX vectors containing genes that encode a pancreatic tumor-associated

XX antigen (PTAA).

XX Example 4; SEQ ID NO 3; 91pp; English.

XX The invention relates to a method of inducing an immunological response

XX against a malignant pancreatic cell in an individual, which comprises:

XX selecting an individual having malignant pancreatic cells or at risk for

XX developing such a pancreatic tumor, administering to the individual a

XX first vector containing a first gene, or its antigenic portion, that

XX encodes a pancreatic tumor-associated antigen (PTAA), and at regular

XX intervals thereafter administering at least a second vector containing a

XX gene encoding PTAA or its antigenic portion, where if carcinoembryonic

XX antigen (CEA) or mucin 1 (MUC-1) or its antigenic portion or modified

XX version is the PTAA, there must be a second PTAA present. The method

XX further comprises administering granulocyte-macrophage colony stimulating

XX factor (GM-CSF) or at least one co-stimulatory molecule. The co-

XX stimulatory molecule is administered as a gene contained within the same

XX or differing vector as the vector containing gene encoding the PTAA.

XX PTAA, or its antigenic portion is contained in a poxvirus vector, such as

XX vaccinia. The PTAA may be a mucin selected from MUC-1, MUC-2, MUC-3, MUC-

CC 4, MUC-SAC, MUC-5B, MUC-6, MUC-7, MUC-11, MUC-12, and their antigenic

CC portions and modified versions. The modified version is wobbled-MUC-1.

CC The method is useful for treating individuals at risk of developing or

CC suffering from pancreatic cancer. The present sequence represents DNA

CC encoding wobbled-CEA (wCEA).

XX SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores: 8.03 Length: 2106

Score: 45.00 Matches: 9

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 14 Gaps: 0

US-10-725-373-2 (1-9) x AEA11047 (1-2106)

QY 1 TyrLeuSerGlyAlaAspLeuAnLeu 9

Db 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 11

AEA12740

ID AEA12740 standard; DNA; 2106 BP.

XX AC AEA12740;

XX DT 28-JUL-2005 (first entry)

XX DE Wobbled CEA DNA.

XX KW immunogenicity; breast tumor; Cytostatic; neoplasm; ds; gene; CEA.

XX OS Unidentified.

XX FH Key

XX CDS

XX Location/Qualifiers

XX 988..2106

XX /*tag= a

XX /product= "wCEA(6D)"

XX /transl_except= (pos:1825..1827,aa:Asn)

XX PN WO2005046614-A2.

XX PD 26-MAY-2005.

XX PF 12-NOV-2004; 2004WO-US037810.

XX PR 12-NOV-2003; 2003US-0519427P.

XX PA (THER-) THERION BIOLOGICS CORP.

XX PI Panicali DL, Mazzara GP, Gritz LR;

XX DR WPI; 2005-386205/39.

XX P-PSDB; AEA12741.

XX Inducing an immunological response against a cell expressing a breast

XX cancer associated antigen in a human, useful for treating breast cancer,

XX by administering vectors containing genes that encode a breast cancer

XX associated antigen.

XX Example 4; SEQ ID NO 3; 99pp; English.

XX The invention relates to a method of inducing an immunological response

XX against a cell expressing a breast cancer associated antigen in a human

XX comprises administering vectors containing genes or its antigenic portion

XX that encode a breast cancer associated antigen. The method is useful for

XX treating individuals suffering from breast cancer. The present sequence

XX represents wobbled CEA DNA.

XX SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.03 Length: 2106
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-2 (1-9) x AEA12740 (1-2106)

QY 1 TyrLeuSerGlyAlaAapLeuAenLeu 9
|||
Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 12

AEBS0768

ID AEB30768 standard; DNA; 2106 BP.

XX

AC AEB30768;

XX 06-OCT-2005 (first entry)

XX Multi-antigen construct CEA-CAP1-6D-1,2.

DE cytostatic; immunostimulant; vaccine; expression; vector; immunotherapy;
KW cancer; neoplasm; cellular transport; ds.
KW Unidentified.

OS WO2005068640-A2.

XX 28-JUL-2005.

PN 23-DEC-2004; 2004WO-US042980.

XX 23-DEC-2003; 2003US-0532205P.

PF (AVET) AVENTIS PASTEUR INC.

XX (THER-) THERION BIOLOGICS INC.

PA Parrington M, Berinstitute N, Tartaglia JT, Panicali D, Gritz L;

XX WPI; 2005-522835/53.

XX Novel expression vector comprising nucleic acid sequences encoding

PT modified KSA, useful for immunizing host, and for preventing or treating

PT cancer.

XX Example 2; Fig 2; 64pp; English.

PS The invention describes an expression vector (I) useful for immunizing a

XX host, comprising nucleic acid sequences encoding modified KSA. Also

CC described are: a composition (CI) comprising (I) in a carrier, where (I)

CC molecule comprising the modified KSA coding sequence of a fully defined

CC and encoding modified KSA having a fully defined approximately 314 amino

CC acid sequence given in the specification; and an isolated DNA molecule

CC comprising carcinoembryonic antigen (CEA), p53, and modified KSA coding

CC sequences, where the CEA sequence is CEA-CAP1-6D-1,2 having a fully

CC defined approximately 2106 nucleotide sequence given in the

CC specification, the p53 sequence has a sequence as given in the

CC CI is also disclosed. (I) is useful for immunizing a host. (I) is useful

CC for preventing or treating cancer, which involves administering (I) to a

CC host, where (I) comprises nucleic acid sequences encoding modified KSA.

CC This sequence represents CEA-CAP1-6D-1,2, a modified version of AEB30767

CC used in multi-antigen construct vcp2086.

XX Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.03 Length: 2106

Score: 45.00

Percent Similarity: 100.00%

Matches: 9

Conservative: 0

Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-2 (1-9) x AEB30768 (1-2106)

QY 1 TyrLeuSerGlyAlaAapLeuAenLeu 9

|||
Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 13

AEBS0767

ID AEB30767 standard; DNA; 2106 BP.

XX

AC AEB30767;

XX 06-OCT-2005 (first entry)

XX Multi-antigen construct CEA-CAP1-6D.

DE cytostatic; immunostimulant; vaccine; expression; vector; immunotherapy;

XX cancer; neoplasm; cellular transport; ds.

XX Unidentified.

OS WO2005068640-A2.

XX 28-JUL-2005.

PN 23-DEC-2004; 2004WO-US042980.

XX 23-DEC-2003; 2003US-0532205P.

PF (AVET) AVENTIS PASTEUR INC.

XX (THER-) THERION BIOLOGICS INC.

PA Parrington M, Berinstitute N, Tartaglia JT, Panicali D, Gritz L;

XX WPI; 2005-522835/53.

XX Novel expression vector comprising nucleic acid sequences encoding

PT modified KSA, useful for immunizing host, and for preventing or treating

PT cancer.

XX Example 2; Fig 2; 64pp; English.

PS The invention describes an expression vector (I) useful for immunizing a

XX host, comprising nucleic acid sequences encoding modified KSA. Also

CC described are: a composition (CI) comprising (I) in a carrier, where (I)

CC molecule comprising the modified KSA coding sequence of a fully defined

CC and encoding modified KSA having a fully defined approximately 314 amino

CC acid sequence given in the specification; and an isolated DNA molecule

CC comprising carcinoembryonic antigen (CEA), p53, and modified KSA coding

CC sequences, where the CEA sequence is CEA-CAP1-6D-1,2 having a fully

CC defined approximately 2106 nucleotide sequence given in the

CC specification, the p53 sequence has a sequence as given in the

CC CI is also disclosed. (I) is useful for immunizing a host. (I) is useful

CC for preventing or treating cancer, which involves administering (I) to a

CC host, where (I) comprises nucleic acid sequences encoding modified KSA.

CC This sequence represents CEA-CAP1-6D-1,2 which is modified as shown in

CC AEB30768 for inclusion in multi-antigen construct vcp2086.

XX Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.03 Length: 2106

Score: 45.00

Percent Similarity: 100.00%

Matches: 9

Conservative: 0

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-2 (1-9) x AEB30767 (1-2106)
QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
Db 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 14
AAI72490/c
ID AAI72490 standard; DNA; 7958 BP.
XX AC AAI72490;
XX DT 16-MAY-2002 (first entry)
XX DE H6-promoter human CEAmoD/42K-promoted B7.1 insertion cassette.
XX KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
XX KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
XX KW prostate; cancer; gene; ds.
XX OS Homo sapiens.
XX OS Synthetic.
XX OS Chimeric.
XX FH Key Location/Qualifiers
FT misc_feature 423..827
FT /*tag= a
FT /note= "ALVAC's C5 locus left flanking arm"
FT CDS complement(903..3008)
FT /*tag= b
FT /product= "CEA agonist peptide"
FT promoter complement(3009..3132)
FT /*tag= c
FT /label= Vaccinia_H6_promoter
FT 3210..3275
FT /*tag= d
FT /label= 42K_promoter
FT 3276..4142
FT CDS /*tag= e
FT /*product= "Human B7.1"
FT misc_feature 4184..5722
FT /*tag= f
FT /note= "ALVAC's C5 locus right flanking arm#"
XX WO200210379-A2.
XX PD 07-FEB-2002.
XX PD 27-JUL-2001; 2001WO-CA001092.
XX PR 31-JUL-2000; 2000US-0222043P.
XX PA (AVET ) AVENTIS PASTEUR LTD.
XX PA (THER-) THERION BIOLOGICS.
XX PA (USSH ) US NAT CANCER INST.
XX PI Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
XX PI Schlom J;
XX DR WPI; 2002-206189/26.
XX PT Carcinoembryonic antigen agonist polypeptide for inducing an immune
XX PT response in animal against antigen and for inhibiting an epitope antigen
XX PT expressing carcinoma cell, comprises a modified antigen epitope.
XX PS Example 3; Fig 4; 69pp; English.
XX CC This sequence represents ALVAC(2)-CEAmoD/hb7.1. This is a coding sequence
XX CC containing the H6 promoted modified carcinoembryonic antigen (CEA)

```

```

CC agonist polypeptide of the invention. The CEA agonist contains a modified
CC CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
CC increase its immunogenicity. The CEA agonist polypeptide of the
CC invention, or DNA encoding it, are useful for: (i) inducing an immune
CC response in an animal directed against a CEA protein or fragment, CEA
CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or
CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer
XX
XX Sequence 7958 BP; 2096 A; 1720 C; 1858 G; 2284 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 36 Length: 7958
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-2 (1-9) x AAI72490 (1-7958)
QY 1 TyrLeuSerGlyAlaAspLeuAsnLeu 9
Db 1199 TACCTTTCGGAGCGGACCTCAACCTC 1173

RESULT 15
AEB30764/c
ID AEB30764 standard; DNA; 8210 BP.
XX AC AEB30764;
XX DT 06-OCT-2005 (first entry)
XX DE ALVAC donor plasmid pNC5LSLSPCEAp53.
XX KW cytostatic; immunostimulant; vaccine; expression; vector; immunotherapy;
XX KW cancer; neoplasm; cellular transport; pNC5LSLSPCEAp53; plasmid; ds;
XX KW circular.
XX OS Synthetic.
XX OS Unidentified.
XX FH Key Location/Qualifiers
FT CDS 3187..4368
FT /*tag= a
FT /product= "p53"
XX WO2005068640-A2.
XX PD 28-JUL-2005.
XX PR 23-DEC-2004; 2004WO-US042980.
XX PR 23-DEC-2003; 2003US-0532205P.
XX PA (AVET ) AVENTIS PASTEUR INC.
XX PA (THER-) THERION BIOLOGICS INC.
XX PI Farrington M, Berinstein N, Tartaglia JT, Panicali D, Gritz L;
XX DR WPI; 2005-522835/53.
XX DR P-PSDB; AEB30766.
XX PT Novel expression vector comprising nucleic acid sequences encoding
XX PT modified KSA, useful for immunizing host, and for preventing or treating
XX PT cancer.
XX PS Example 1; Fig 1; 64pp; English.
XX CC The invention describes an expression vector (I) useful for immunizing a

```


CC host, comprising nucleic acid sequences encoding modified KSA. Also
 CC described are: a composition (C1) comprising (I) in a carrier, where (I)
 CC comprises nucleic acid sequences encoding modified KSA; an isolated DNA
 CC molecule comprising the modified KSA coding sequence of a fully defined
 CC approximately 945 nucleotide sequence (S1) given in the specification,
 CC and encoding modified KSA having a fully defined approximately 314 amino
 CC acid sequence given in the specification; and an isolated DNA molecule
 CC comprising carcinoembryonic antigen (CEA), p53, and modified KSA coding
 CC sequences, where the CEA sequence is CEA-CAP1-6D-1,2 having a fully
 CC defined approximately 2106 nucleotide sequence given in the
 CC specification, the p53 sequence has a sequence as given in the
 CC specification, and the modified KSA sequence has (S1). A kit comprising
 CC C1 is also disclosed. (I) is useful for immunizing a host. (I) is useful
 CC for preventing or treating cancer, which involves administering (I) to a
 CC host, where (I) comprises nucleic acid sequences encoding modified KSA.
 CC This sequence represents a plasmid pNC51SPCEAP53 encoding CEA and p53
 CC polypeptides and used in the creation of multi-antigen construct vcp2086.
 CC Note: This sequence does not appear to encode AEB30765 as suggested in
 CC figure 1.

XX
 SQ Sequence 8210 BP; 2135 A; 1864 C; 1925 G; 2286 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	37.3	Length:	8210
Score:	45.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-725-373-2 (1-9) x AEB30764 (1-8210)

QY	1	TyrLeuSerGlyAlaLeuLeuLeu	9
DB	1182	TACCTTTCGGAGCGACCTCAACCTC	1156

Search completed: December 6, 2005, 16:29:49
 Job time : 374.25 secs


```

Key source Location/Qualifiers
FT source 1..27
FT /organism="Homo sapiens (human)"
FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0.00469 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-725-373-3 (1-9) x BD131677 (1-27)
QY 1 TyrLeuSerGlyAlaAspIleAenLeu 9
DB 1 TACCTTCGGGAGCGGACATCAACCTC 27
RESULT 2
LOCUS CS089179 27 bp DNA linear PAT 25-MAY-2005
DEFINITION Sequence 8 from Patent EP1447414.
ACCESSION CS089179
VERSION CS089179.1 GI:66714458
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Schlom,J., Salazar,M.E. and Zarembo,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: EP 1447414-A 8 18-AUG-2004;
DEPARTMENT OF Health and Human Services (US)
FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0.00469 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-725-373-3 (1-9) x CS089179 (1-27)
QY 1 TyrLeuSerGlyAlaAspIleAenLeu 9
DB 1 TACCTTCGGGAGCGGACATCAACCTC 27
RESULT 3
AR560606
LOCUS AR560606 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 8 from patent US 6756038.
ACCESSION AR560606
VERSION AR560606.1 GI:53972927
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 27)
Schlom,J., Barzaga,E. and Zarembo,S.
Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
Patent: US 6756038-A 8 29-JUN-2004;
The United States of America as represented by the Department of
Health and Human Services; Washington, DC;
WOX;
FEATURES
source 1..27
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0.00469 Length: 27
Score: 45.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-725-373-3 (1-9) x AR560606 (1-27)
QY 1 TyrLeuSerGlyAlaAspIleAenLeu 9
DB 1 TACCTTCGGGAGCGGACATCAACCTC 27
RESULT 4
LOCUS BD131676 27 bp DNA linear PAT 18-SEP-2002
DEFINITION Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
ACCESSION BD131676
VERSION BD131676.1 GI:23226621
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zarembo,S.
TITLE Carcinoembryonic antigen (CEA) agonist and antagonist peptides
JOURNAL Patent: JP 2002500002-A 2 08-JAN-2002;
THE UNITED STATES OF AMERICA
COMMENT OS Homo sapiens (human)
PN JP 2002500002-A/2
PD 08-JAN-2002 JP 2000516030
PF 22-SEP-1998 JP 2000516030
PR 10-OCT-1997 US 60/061589
PI JEFFREY SCHLOM,ELENE BARZAGA,SAM ZAREMBA
PC C12N15/09,A61K38/00,A61K45/00,A61K48/00,A61P35/00,A61P37/02,
A61P43/00
PC C07K14/705,C07K16/28,C12N15/10,C12N15/00,A61K37/02,C12N5/00 CC
Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH
Key Location/Qualifiers
FT source 1..27
/organism="Homo sapiens (human)"
FEATURES
source 1..27
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0.0154 Length: 27
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0
US-10-725-373-3 (1-9) x BD131676 (1-27)
```

```

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db 1 TACCTTTTCGGGAGCGGACCTCAACCTC 27

RESULT 5
LOCUS CS089178 27 bp DNA linear PAT 25-MAY-2005
DEFINITION Sequence 7 from Patent EP1447414.
ACCESSION CS089178.1 GI:66714457
VERSION CS089178.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Schlom,J., Salazar,M.E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: EP 1447414-A 7 18-AUG-2004;
Department of Health and Human Services (US)
FEATURES
source
1..27
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.0154 Length: 27
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x CS089178 (1-27)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db 1 TACCTTTTCGGGAGCGGACCTCAACCTC 27

RESULT 6
LOCUS AR560605 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 7 from patent US 6756038.
ACCESSION AR560605
VERSION AR560605.1 GI:53972926
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 27)
AUTHORS Schlom,J., Barzaga,E. and Zaremba,S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: US 6756038-A 7 29-JUN-2004;
The United States of America as represented by the Department of
Health and Human Services; Washington, DC;
WOX;

FEATURES
source
1..27
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0.0154 Length: 27
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x AR560605 (1-27)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db 1 TACCTTTTCGGGAGCGGACCTCAACCTC 27

RESULT 7
LOCUS AX133657 2106 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 111 from Patent WO0130382.
ACCESSION AX133657
VERSION AX133657.1 GI:14139699
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berinstein,N., Tartaglia,J., Molineon,P. and Barber,B.
TITLE Method of inducing and/or enhancing an immune response to tumor
antigens
JOURNAL Patent: WO 0130382-A 111 03-MAY-2001;
Aventis Pasteur Limited (CA)
FEATURES
Location/Qualifiers
1..2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1..2106
/note="unnamed protein product; modified CEA"

ORIGIN
Alignment Scores:
Pred. No.: 2.86 Length: 2106
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x AX133657 (1-2106)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db 1810 TACCTTTTCGGGAGCGGACCTCAACCTC 1836

RESULT 8
LOCUS AX192349 2106 bp DNA linear PAT 15-AUG-2001
DEFINITION Sequence 3 from Patent WO0149317.
ACCESSION AX192349
VERSION AX192349.1 GI:15210326
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1

```

AUTHORS Entage,P., Barber,B.H., Sambhara,S. and Sia,C.D.
TITLE Enhancing the immune response to an antigen by presensitizing with
an inducing agent prior to immunizing with the inducing agent and
the antigen
JOURNAL Patent: WO 0149317-A 3 12-JUL-2001;
Aventis Pasteur Limited (CA)
FEATURES
source
1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1. .2106
/note="unnamed protein product; modified CEA"

CDS

/codon_start=1
/transl_table=1
/protein_id="CAC51319.1"
/db_xref="GI:15210327"
/translation="MESPAPAPHCIPQWRLLLTASLTFFWNPPTAKLTISTPFP
VAEGKEVLVHNLPHQFGYSGYKGERVDGNRIIGYVIGTQQATPGPAYGRELIY
PNASLLIQNIQDTGFTYTHVTKSLVNEEATGQFRVPELPKPSISNNKPVEDK
DAVAFCEPETQDATVLMVNNQSLPVSRLSGNGNRTLFPNTRNDTASVKCEIQ
NPVSARRSDVILNVLYGDAPTISPLNTSVSGENLNLSCHAASNPAPQYKWFVNGT
FQQTQELFPNITVNNSSGYTCQAHSITGLNRTVTITTYVEPKPPIITNNSNPV
EDEDVAUTCPEIQNTLVYLMVNNQSLPVSRLSGNDRNLTLISVTRNDVGPYEC
GIONELSDVHSDPEVILNVLYGDDPTISPSYTYRPGVNLNLSCHAASNPAPQYSLI
DGNIOQHTOELFISNITEKNGLYTCQAHSNGHRTVTKTITVSAELPKPSISNN
SKPVEDKDAVAFCEPEAQNTVYLMVNNQSLPVSRLSGNRTLTLPNTRNDAR
AYVCGIQNSVSNRSDPTVLDVLYGDDPTIISPDDSSVLSGADLNLSCHAASNPSPQY
SWRINGIPQOHTQVLFIAKITPNNGNYACFVSNLATGRNNSIVKSIIVSASGTSFGL
SAGATVGINIGVLGVGVALI"

ORIGIN

Alignment Scores: 2.86 Length: 2106
Pred. No.: 8
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x AX192349 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAenLeu 9
Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836
RESULT 9
AX393888
LOCUS AX393888 2106 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 2 from Patent WO0210379.
ACCESSION AX393888
VERSION AX393888.1 GI:19701852
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Berinstein,N., Tartaglia,J., Tine,J.A., Panicali,D.L., Gritz,L. and Schlom,J.
TITLE Modified cea and uses thereof
JOURNAL Patent: WO 0210379-A 2 07-FEB-2002;
Aventis Pasteur Limited (CA) ; Therion Biologics (US) ; National Cancer Institute (US)
FEATURES
source
1. .2106
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Cea modified polypeptide"

ORIGIN

Alignment Scores: 2.86 Length: 2106
Pred. No.: 8

Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x AX393888 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAenLeu 9
Db 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 10
AC154826/c

LOCUS AC154826 150028 bp DNA linear ROD 20-JUL-2005
DEFINITION Mus musculus BAC clone RP23-72D18 from chromosome 9, complete
sequence.

ACCESSION AC154826 AC136727
VERSION AC154826.2 GI:71037564
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 150028)
AUTHORS Zheng,X. and Haglund,K.

TITLE The sequence of Mus musculus BAC clone RP23-72D18

JOURNAL Unpublished (2001)

REFERENCE 2 (bases 1 to 150028)

AUTHORS Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park

Way, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 150028)

AUTHORS Wilson,R.K.

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2005) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT On Jul 20, 2005 this sequence version replaced gi:56900452.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@watson.wustl.edu

----- Summary Statistics

Center project name: M_BA0072D18

Drafting center: WIBR

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e. phred quality
>=30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone, fosmid clone or direct clone walk sequence.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to obtain the consensus sequence. The
assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous
human standard. Specifically, standards for regions of low sequence
complexity (such as dinucleotide repeats and small unit tandem
repeats) have been relaxed. These regions are very prevalent in the
mouse genome, and the return on extended finishing efforts is
minimal.

If a sequence meets the criteria of the above statement, it needs
no comments or tags. If the criteria are not met, such as ambiguous
bases, then the region is duly annotated.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,

Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The BAC Library has been constructed by Kazutoyo Oeegawa and Minako Tatenko in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES

```

source
  1. .150028
    /location="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="9"
    /clone_lib="RP41-23"
    /clone="RP23-72D18"
    /size="57374"
    /notes="Unresolved simple sequence repeat."
misc_feature
  91044..91104
    /notes="Sequence derived from PCR product of project DNA"
  91395
    /notes="Unresolved bases"
  102401..102416
    /notes="Sequence derived from one plasmid subclone."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      862      Length:      150028
Score:          42.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 88.89%      Mismatches: 0
Query Match:      93.33%      Indels: 0
DB:              9      Gaps: 0

```

US-10-725-373-3 (1-9) x AC154826 (1-150028)

QY 1 TyLeuSerGlyAlaAepIleAsnLeu 9

```

|||||
Db 80666 TATCTAGTGGGCTGCATCACCTT 80640

```

```

RESULT 11
AC149846/c
LOCUS
DEFINITION
  AC149846 185385 bp DNA linear HTG 06-AUG-2004
  Papio anubis clone RP41-153B23, WORKING DRAFT SEQUENCE, 4 ordered
  pieces.
ACCESSION
  AC149846
VERSION
  AC149846.2 GI:51011162
KEYWORDS
  HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE
  Papio anubis (olive baboon)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Cercopithecoidea; Cercopithecinae; Papio.

```

```

REFERENCE
  1 (bases 1 to 185385)
  Antonellis A., Avele K., Benjamin B., Blakesley R.W.,
  Bouffard G.G., Brinkley C., Brooks S., Chu G., Coleman B.,
  Coleman H., Daki N., Engle J., Guan X., Gupta J., Haghghi P.,
  Han J., Hansen N., Ho S.-L., Hu P., Hurle B., Idol J.R., Jones C.,
  Karlins E., Kim H., Kwong P., Laric P., Larson S., Lee-Lin S.-Q.,
  Legaspi R., Madden M., Maduro Q.L., Maduro V.B., Margulies E.H.,
  Masillo C., Maskeri B., McDowell J., Mullikin J.C., Paquirigan C.,
  Park M., Portnoy M.E., Prasad A., Puri O., Reddix-Dugue N.,
  Schandler K., Schueler M.G., Shah K., Sison C., Stantropop S.,
  Thomas J.W., Thomas P.J., Tsipouri V., Vogt J.L., Wetherby K.D.,
  Young A. and Green E.D.

```

```

AUTHORS
  NISC Comparative Sequencing Initiative
  Unpublished
  2 (bases 1 to 185385)

```

```

TITLE
JOURNAL
REFERENCE

```

AUTHORS

Direct Submission
Submitted (23-JUN-2004) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

REFERENCE

3 (bases 1 to 185385)

AUTHORS

Direct Submission
Submitted (06-AUG-2004) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Aug 6, 2004 this sequence version replaced gi:49065688.

COMMENT

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc.zoo@hgri.nih.gov

----- Project Information

Center project name: hpx

Center clone name: 153B23

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 183688 bases at least Q40

Consensus quality: 184188 bases at least Q30

Consensus quality: 184625 bases at least Q20

Insert size: 180000; agarose-fp

Insert size: 185085; sum-of-contigs

Quality coverage: 10.31x in Q20 bases; agarose-fp

Quality coverage: 10.03x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 104105: contig of 104105 bp in length

* 104106 104205: gap of unknown length

* 104206 109148: contig of 4943 bp in length

* 109149 109248: gap of unknown length

* 109249 130407: contig of 21159 bp in length

* 130408 130507: gap of unknown length

* 130508 185385: contig of 54878 bp in length.

FEATURES

source

1. 185385

/organism="Papio anubis"

/mol_type="genomic DNA"

/db_xref="taxon:9555"

/clone_lib="RP41-153B23"

/note="TAC resource: <http://bacpac.chori.org/>"

misc_feature

1. 104105

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

misc_feature

1. 57045

/note="clone overlaps with GenBank Accession Number

AC150306 clone RP41-49E22 (center project name hgg) "
104106. .104205

mays Anthocyanin Pathway Have Distinct Phenotypes Relative to the Functionally Similar TRANSPARENT TESTA GLABRA1 Gene in Arabidopsis thaliana

JOURNAL PUBMED
REFERENCE 14742877
AUTHORS 2 (bases 1 to 8239)
TITLE Carey,C.C., Chandler,V.L. and Strahle,J.
JOURNAL Direct Submission
SUBMITTED Submitted (28-MAY-2002) Plant Sciences, University of Arizona, 303 Forbes Building, Tucson, AZ 85721, USA
FEATURES Location/Qualifiers
source 1..8239
organism="Zea mays"
mol_type="genomic DNA"
cultivar="B73"
db_xref="taxon:4577"
chromosome="5"
map="within 3.5 cm of phi087 SSR marker (phi087 is at coordinate 103.3 on chromosome 5 on Pioneer Composite 1999 map)"
clone="ZMMBB0124P19"
c3320..7854
gene /gene="paci"
mRNA join(<3320..4385,7403..7854)
CDS /product="anthocyanin biosynthetic gene regulator PAC1"
3320..4381
/gene="paci"
note=WD40 repeat protein; required for transcriptional activation of anthocyanin biosynthetic genes; possible transcriptional regulator
codon_start=1
/product="anthocyanin biosynthetic gene regulator PAC1"
protein_id="AAW76742.1"
db_xref="GI:37544703"
translation="MDFPKPPSSVASSSGPETPNHPTCELPHSIYALAFSPVPVLASGFLDLHNRLSLGSFDPVRPAELSRFPALPSFDHPYPPTKLQFNPRAAAPSLALA SSADTLRIWHPTLDDLSDTAPAEPLRSVLDNRKASSEFCAPLTFSFDNVEPRRIGTA SIDTTCTWVIDRGVETQTLLAHDKAVHDIAWGAGVFASVDGSGVRVFDLRDKRHES TIVYESRPDTPLLRLAWNRSDLRYMAALLMDSSAVVLDIRAPGVPAELHRRACAA NAWAKAPATHLCSGDDGOALLIWELPEATAA VPAEGIDPVLYDGAGEINQLQMAA AHPDWAMATFAENKVQLLRV"
unsure 3772
/gene="paci"
note="Sequencing ambiguity; may encode Glu or Asp"
join(4382..4385,7403..7854)
/gene="paci"

Alignment Scores:
Pred. No.: 48.3 Length: 8239
Score: 41.00 Matches: 7
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 77.78% Mismatches: 0
Query Match: 91.11% Indels: 0
DB: 15 Gaps: 0

US-10-725-373-3 (1-9) x AY115485 (1-8239)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
|||||...:::|
Db 2313 TATTTAAGTGCGTCGTGAATTA 2339
|||...:::|

RESULT 14
LOCUS AL161654/c
DEFINITION Human DNA sequence from clone RP11-59G22 on chromosome 10 Contains part of a novel gene (KIAA1136) (FLJ37801) and a CpG island, complete sequence.
ACCESSION AL161654
VERSION AL161654.14 GI:14970795
KEYWORDS HTG; CpG island; FLJ37801; KIAA1136.
SOURCE Homo sapiens (human)


```
/note="match: proteins: Q9GKT3"
/codon_start=1
/product="novel protein"
/protein_id="CAI16106.1"
/db_xref="GI:55957959"
/db_xref="GOA:Q5T848"
/db_xref="InterPro:IPR000337"
/db_xref="UniProt/TREMBL:Q5T848"
/translation="MGAMAYPLLLCLLAQLGLGAVGASRD PQGRDPS PRRTPKGKP
HAQQGRASASDSSAPWSRSTDTGTLIAQKLAEEVPMDSVLYLTGDSHQLKRCNSGR
YELALPGKWPALASAPSLHRLADLTHTATFLNVLMSQNSREQLQDDLDWQAL
VNSLGEPSISRAAITFTSDLSAPAPQVFLQATREESRIILLQDLSAPHLANATL
ETWFHGLRPHILRRGPNQGRGLGSHKDFKWSPPYLECENG
YKPMPLTSSAIYGLQPNLVEFGVMKVDINLOKVIDQCSSGDFGTHKCHLN
SECMPTKIGFVGLGVECEIKAGFVHPGVLPVNNFRPGPDHISGSTKDVSEAVVC
LPCRGCPACDSCFVQEDKILRLAIIISFQALCMLLDFVSMLVVYHFRKAKSTRAS
GILLETILFGSLLYFFPVILYFEPSTPCTILLRRLGLGATVGTVTLKHLRVLK
VELSTAQIPMYTQGRVMRLAVILLVFWFLIGWTSSVCQNLKQISLIQGGKTS
HLIFNMCLIDWDYMTAVAEFLFLGVLVYAVRTVPSAFHEPRYMAVAVNELIIS
ALFHTIRFVLASRLQSDMLMLYFAHTHTVTVTIGLLI PKFSSNNPRDDIATEA
VEDELDMRSGSYLNSINSAMSEHSLDPEDIRDELKLYAOLEYKPKMTIANNPHL
QXKRSCKGLGRSIRWRIETPVSQCSKEDKEDAGTGTAKGTALIRKNPESGNN
TKSKEETLKNRVFLSKSHSTYDHRDQTERSSSLPTESQBEETTESLTSKCK
LTQKLEBSEASTESVPLVCKSASAHNLSSEKKTGHPRTSMQKSLSVIASAKETL
GLAGKTQTAGVEERTSKQPLPKDKETNRNHSNNTETKDPAPQNSNPAEPRKPKQ
SGIMKQORVNPFTANSDLNPGTOMKDNFDIGEVCPWEVYDLPGPVRESKVQKHVS
IVASMEKNPTSLKESHKPKAAEVCCQSNKRIDKAEVCLMESQGSILDEKLL
ISKTPVLERAKEENGQPPRAANVCAGQSEELPPKAVASKTENENLQIGHKEKTS
SEENVGSGNSNNTPQPTSRAEVCPWEFETPAQPNAGRSVALPASSALSANKIAGP
RKEEWDSPFKV"
gene join(110700..110868,111161..111229,119941..120343)
/gene="RP11-59G22.1"
/locus_tag="RP11-59G22.1-002"
join(110700..110868,111161..111229,119941..120343)
/gene="RP11-59G22.1"
/locus_tag="RP11-59G22.1-002"
/product="novel protein"
/note="match: ESTs: BG687462.1"
127080
/note="Clone_right_end: RP13-236A4"
150489
/note="Clone_right_end: RP11-561H23"
159231
/note="Clone_right_end: RP11-59G22"
ORIGIN
Alignment Scores:
Pred. No.: 1.68e+03 Length: 159231
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 91.11% Indels: 0
DB: 8 Gaps: 0
US-10-725-373-3 (1-9) x AL161654 (1-159231)
QY 1 TyrLeuSerGlyAlaAsp1LeuLeu 9
|||||
Db .137882 TATCTGAGTGGTGACAGATAAACTTG 137856
RESULT 15
AL591477/c AL591477 168608 bp DNA linear HTG 19-DEC-2001
LOCUS Homo sapiens chromosome 10 clone RP13-112H19, WORKING DRAFT
DEFINITION SEQUENCE, 3 unordered pieces.
ACCESSION AL591477
VERSION AL591477.2 GI:17973980
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
```

```
REFERENCE
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 20, 2001 this sequence version replaced gi:14141526.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: b1112H19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 167544 bases at least Q40
Consensus quality: 167716 bases at least Q30
Consensus quality: 167994 bases at least Q20
Insert size: 168408; sum-of-contigs
Insert size: 152446; agarose-fp
Quality coverage: 9.85x in Q20 bases; sum-of-contigs Quality
coverage: 10.88x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 35137: contig of 35137 bp in length
* 35138 35237: gap of 100 bp
* 35238 60541: contig of 25304 bp in length
* 60542 60641: gap of 100 bp
* 60642 168608: contig of 107967 bp in length.
FEATURES
source
1..168608
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RP13-112H19"
/clone_end:T7
/notes="assembly_fragment:00872"
1..35137
vector_side:left"
vector_side:right"
35238..60541
/notes="assembly_fragment:01936"
60642..168608
/notes="assembly_fragment:00906"
clone_end:SP6
vector_side:right"
misc_feature
1..35137
/notes="assembly_fragment:00872"
clone_end:T7
vector_side:left"
misc_feature
35238..60541
/notes="assembly_fragment:01936"
60642..168608
/notes="assembly_fragment:00906"
clone_end:SP6
vector_side:right"
ORIGIN
Alignment Scores:
Pred. No.: 1.8e+03 Length: 168608
Score: 41.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 91.11% Indels: 0
DB: 14 Gaps: 0
US-10-725-373-3 (1-9) x AL591477 (1-168608)
QY 1 TyrLeuSerGlyAlaAsp1LeuLeu 9
|||||
Db 36233 TATCTGAGTGGTGACAGATAAACTTG 36207
```

Search completed: December 6, 2005, 19:51:35
Job time : 3055.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 10:17:58 ; Search time 367.25 Seconds
(without alignments)
163.328 Million cell updates/sec

Title: US-10-725-373-3
Perfect score: 45
Sequence: 1 YLSGADINL 9

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US10725373/runat_01122005_114444_21420/app.query.fasta_1.796
-DB=N_Geneseq -QMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=JUS10725373 @CGN_1_1244 @runat_01122005_114444_21420 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_21:*

1:	Geneseqn1980s:*
2:	Geneseqn1990s:*
3:	Geneseqn2000s:*
4:	Geneseqn2001as:*
5:	Geneseqn2001bs:*
6:	Geneseqn2002as:*
7:	Geneseqn2002bs:*
8:	Geneseqn2003as:*
9:	Geneseqn2003bs:*
10:	Geneseqn2003cs:*
11:	Geneseqn2003ds:*
12:	Geneseqn2004as:*
13:	Geneseqn2004bs:*
14:	Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	27	2	Aax56259 Carcinoem
2	43	95.6	27	2	Aax56258 Carcinoem
3	43	95.6	2105	6	Aai72497 CEA agoni
4	43	95.6	2106	4	Aah20121 Modified

5	43	95.6	2106	5	AAD07347	Aad07347 Modified
6	43	95.6	2106	6	AAI72489	Aai72489 CEA agoni
7	43	95.6	2106	10	ADE13860	Adel13860 CEA-CAP6D
8	43	95.6	2106	10	ADE13861	Adel13861 CEA (6D) -1
9	43	95.6	2106	14	ADZ58977	Adz58977 Novel CEA
10	43	95.6	2106	14	ADZ58978	Adz58978 Novel CEA
11	43	95.6	2106	14	AEa11047	Aea11047 DNA encod
12	43	95.6	2106	14	AEa11047	Aea11047 Wobbled C
13	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
14	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
15	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
16	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
17	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
18	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
19	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
20	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
21	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
22	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
23	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
24	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
25	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
26	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
27	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
28	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
29	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
30	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
31	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
32	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
33	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
34	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
35	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
36	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
37	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
38	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
39	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
40	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
41	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
42	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
43	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
44	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant
45	43	95.6	2106	14	AEa12740	Aea12740 Multi-ant

ALIGNMENTS

RESULT 1	
AAX56259	
ID AAX56259 standard; DNA; 27 BP.	
XX	
AC AAX56259;	
XX	
DT 20-JUL-1999 (first entry)	
XX	
DE Carcinoembryonic antigen peptide encoding DNA SEQ ID NO:8.	
XX	
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;	
KW immune response; carcinoma; gastrointestinal; breast; pancreatic;	
KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;	
KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.	
XX	
OS Homo sapiens.	
OS Synthetic.	
XX	
FN WO9919478-A1.	
XX	
PD 22-APR-1999.	
XX	
PF 22-SEP-1998; 98WO-US019794.	
XX	
PR 10-OCT-1997; 97US-0061589P.	
XX	
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	

PI Schlom J, Barzaga E, Zaremba S;
 DR WPI; 1999-326544/27.
 PT Peptide agonists and antagonists of carcinoembryonal antigen.
 XX Claim 22; Page 20; 72pp; English.

XX The present invention describes peptides (A) that comprise agonists (Ia) or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are used in vaccines to kill or inhibit carcinoma cells that express CEA or its epitopes, particularly for treating gastrointestinal, breast, pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also be used to proliferate T cells, e.g. from vaccinated subjects, for use in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction to cancer immunotherapy (i.e. to prevent attack on normal but CEA-expressing cells). (Ia) are more active than native sequence (I) and generate a highly specific and systemic anti-CEA response. Cytotoxic T cells generated recognize both (Ia) and native CEA epitopes. The present sequence encodes a specifically claimed example of (Ia)

SQ Sequence 27 BP; 6 A; 9 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.0376 Length: 27
 Score: 45.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-725-373-3 (1-9) x AAX56259 (1-27)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
 DB 1 TACCTTCGGAGCGGACATCAACCTC 27

RESULT 2
 AAX56258
 ID AAX56258 standard; DNA; 27 BP.
 AC AAX56258;
 XX
 DT 20-JUL-1999 (first entry)
 DE Carcinoembryonic antigen peptide agonist encoding DNA SEQ ID NO:7.
 KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
 KW immune response; carcinoma; gastrointestinal; breast; pancreatic;
 KW bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
 KW adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
 XX Homo sapiens.
 OS Synthetic.
 XX WO9919478-A1.
 PN 22-APR-1999.
 XX 22-SEP-1998; 98WO-US019794.
 PF 10-OCT-1997; 97US-0061589P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Schlom J, Barzaga E, Zaremba S;
 PI WPI; 1999-326544/27.
 DR Peptide agonists and antagonists of carcinoembryonal antigen.
 XX Claim 22; Page 20; 72pp; English.

XX The present invention describes peptides (A) that comprise agonists (Ia) or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are used in vaccines to kill or inhibit carcinoma cells that express CEA or its epitopes, particularly for treating gastrointestinal, breast, pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also be used to proliferate T cells, e.g. from vaccinated subjects, for use in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction to cancer immunotherapy (i.e. to prevent attack on normal but CEA-expressing cells). (Ia) are more active than native sequence (I) and generate a highly specific and systemic anti-CEA response. Cytotoxic T cells generated recognize both (Ia) and native CEA epitopes. The present sequence encodes a specifically claimed example of (Ia)

SQ Sequence 27 BP; 5 A; 10 C; 6 G; 6 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.108 Length: 27
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 2 Gaps: 0

US-10-725-373-3 (1-9) x AAX56258 (1-27)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
 DB 1 TACCTTCGGAGCGGACCTCAACCTC 27

RESULT 3
 AAI72497
 ID AAI72497 standard; DNA; 2105 BP.
 AC AAI72497;
 XX 16-MAY-2002 (first entry)
 DE CEA agonist coding sequence #2.
 KW CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
 KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
 KW prostate; cancer; gene; ss.
 XX Synthetic.
 OS Key
 FH Location/Qualifiers
 CDS 1..2105
 FT /tag= a
 FT /product= "CEA agonist polypeptide"
 XX WO200210379-A2.
 PN 07-FEB-2002.
 XX 27-JUL-2001; 2001WO-CA001092.
 PF 31-JUL-2000; 2000US-0222043P.
 XX (AVET) AVENTIS PASTEUR LTD.
 PA (THER-) THERION BIOLOGICS.
 PA (USSH) US NAT CANCER INST.
 XX Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
 PI Schlom J;
 XX WPI; 2002-206189/26.
 DR Carcinoembryonic antigen agonist polypeptide for inducing an immune
 PT response in animal against antigen and for inhibiting an epitope antigen
 PT expressing carcinoma cell, comprises a modified antigen epitope.
 XX

xx The present invention describes a method for inducing an immune response, cc in an animal, to a tumour antigen (Ag) comprising administering Ag, or cc nucleic acid (I) that encodes it, cc to a lymphatic site. Cynomolgus monkeys

XX PS Claim 9; Fig 3; 62pp; English.

XX CC The invention relates to a method of enhancing an immune response against

CC CC tumour-associated antigens (TAAs), such as GP100 and carcinoembryonic

CC CC antigen (CEA) in an animal. The method involves priming of the animal

CC CC with an inducing agent such as tetanus toxoid (TT) or diphtheria toxoid

CC CC (DT), subsequently followed by administration of an inducing agent-

CC CC antigen mixture. The method provides the enhancement or augmentation of

CC CC the immune response to the antigen and/or improves a vaccination protocol

CC CC by allowing use of less antigen. The immunisation of the animal with

CC CC tumour-associated antigen is useful for the prophylactic or therapeutic

CC CC treatment of cancer. The present DNA sequence encodes modified

CC CC carcinoembryonic antigen (CEA) related to the invention

XX SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

XX Alignment Scores:

Pred. No.:	15	Length:	2106
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	5	Gaps:	0

US-10-725-373-3 (1-9) x AAD07347 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9

DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 6

AAI72489

ID AAI72489 standard; DNA; 2106 BP.

XX AC AAI72489;

XX DT 16-MAY-2002 (first entry)

XX DE CEA agonist coding sequence #1.

XX CC CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;

KW carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;

KW prostate; cancer; gene; ds.

XX OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..2106

FT /*tag= a

FT /product= "CEA agonist polypeptide"

XX WO200210379-A2.

PN 07-FEB-2002.

PD 27-JUL-2001; 2001WO-CA001092.

XX 31-JUL-2000; 2000US-0222043P.

XX (AVET) AVENTIS PASTEUR LTD.

PA (THER-) THERION BIOLOGICS.

PA (USSH) US NAT CANCER INST.

XX Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;

PI Schlom J;

XX WPI; 2002-206189/26.

DR P-PSDB; AAB47918.

XX Carcinoembryonic antigen agonist polypeptide for inducing an immune

PT response in animal against antigen and for inhibiting an epitope antigen

PT expressing carcinoma cell, comprises a modified antigen epitope.

XX PS Claim 4; Fig 1; 69pp; English.

XX CC This sequence encodes the carcinoembryonic antigen (CEA) agonist

CC CC polypeptide of the invention. This sequence represents the sequence given

CC CC in the figures in the specification, and it directly encodes the CEA

CC CC agonist polypeptide given in AAB47918. The CEA agonist contains a

CC CC modified CAP-1 epitope of CEA, in which position 6 is modified from Asp

CC CC to Asn to increase its immunogenicity. The CEA agonist polypeptide of the

CC CC invention, or DNA encoding it, are useful for: (i) inducing an immune

CC CC response in an animal directed against a CEA protein or fragment, CEA

CC CC agonist, a CEA epitope, a modified CEA epitope, cells expressing or

CC CC binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope

CC CC expressing carcinoma cell, which is a gastrointestinal, breast,

CC CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a

CC CC patient, hence is useful for manufacture of a medicament for the

CC CC treatment of cancer

XX SQ Sequence 2106 BP; 559 A; 658 C; 442 G; 447 T; 0 U; 0 Other;

XX Alignment Scores:

Pred. No.:	15	Length:	2106
Score:	43.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	95.56%	Indels:	0
DB:	6	Gaps:	0

US-10-725-373-3 (1-9) x AAI72489 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9

DB 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 7

ADE13860

ID ADE13860 standard; DNA; 2106 BP.

XX AC ADE13860;

XX DT 29-JAN-2004 (first entry)

XX CC CEA-CAP6D nucleotide sequence SEQ ID NO:23.

DE carcinoembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;

KW tumour antigen; immunotherapy; gene; ds.

XX OS Unidentified.

XX WO2003085087-A2.

XX PD 16-OCT-2003.

XX 09-APR-2003; 2003WO-US010916.

XX 09-APR-2002; 2002US-0372972P.

XX (AVET) AVENTIS PASTEUR LTD.

PA (THER-) THERION BIOLOGICS INC.

XX Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;

XX WPI; 2003-877029/81.

XX New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-

PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or

PT determining the effectiveness of a chemotherapeutic or other treatment

PT regimen.

XX Example 1; SEQ ID NO 23; 56pp; English.

XX The present invention describes an isolated DNA molecule comprising the

CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see

CC ADE13861), or its fragment. Also described: (1) an expression vector
 CC comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
 CC describes above; (2) a composition comprising the expression vector of
 CC (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
 CC comprising administering to a host the expression vector of (1). CEA(6D)-
 CC 1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
 CC nucleic acid and target polypeptide are useful for diagnosing, preventing
 CC and treating cancer, predicting prognosis, or determining the
 CC effectiveness of a chemotherapeutic or other treatment regimen. The
 CC expression vector may be used for the insertion and expression of CEA(6D)
 CC -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
 CC treatment of cancer. The target polypeptides are useful in generating
 CC antibodies used in screening assays or for immunotherapy. The present
 CC sequence represents the CEA-CAP6D nucleotide sequence, which is given in
 CC comparison with CEA(6D)-1,2 in the exemplification of the present
 CC invention.

XX
 SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-3 (1-9) x ADE13860 (1-2106)

QY 1 TyrluSerGlyAlaAspIleAsnLeu 9
 |||||
 Db 1810 TACCTTCGGGAGCGGACCTCAACCTC 1836

RESULT 8
 ADE13861
 ID ADE13861 standard; DNA; 2106 BP.
 XX
 AC ADE13861;
 XX
 DT 29-JAN-2004 (first entry)
 DE CEA(6D)-1,2 nucleotide sequence SEQ ID NO:24.
 XX
 DE carcinomaembryonic antigen; CEA; CEA(6D)-1; 2; cytostatic; vaccine; cancer;
 KW tumour antigen; immunotherapy; gene; ds.
 XX
 OS Unidentified.

XX WO2003085087-A2.
 XX
 PD 16-OCT-2003.
 XX
 PF 09-APR-2003; 2003WO-US010916.
 XX
 PF 09-APR-2002; 2002US-0372972P.
 PR
 XX (AVET) AVENTIS PASTEUR LTD.
 PA (THER-) THERION BIOLOGICS INC.
 XX
 XX Parrington M, Zhang L, Rovinski B, Gritz LR, Greenhalgh T;
 XX WPI; 2003-877029/81.

XX New isolated DNA molecule comprising the carcinoembryonic antigen (6D)-
 PT 1,2 sequence, useful for diagnosing, preventing and treating cancer, or
 PT determining the effectiveness of a chemotherapeutic or other treatment
 PT regimen.

XX Claim 1; SEQ ID NO 24; 56pp; English.

XX The present invention describes an isolated DNA molecule comprising the
 CC carcinoembryonic antigen (CEA) (6D)-1,2 sequence of 2106 bp (see
 CC ADE13861), or its fragment. Also described: (1) an expression vector

CC comprising the nucleic acid sequence CEA(6D)-1,2, or its fragment
 CC describes above; (2) a composition comprising the expression vector of
 CC (1) in a pharmaceutical carrier; and (3) preventing or treating cancer
 CC comprising administering to a host the expression vector of (1). CEA(6D)-
 CC 1,2 has cytostatic activity, and can be used in vaccines. The CEA(6D)-1,2
 CC nucleic acid and target polypeptide are useful for diagnosing, preventing
 CC and treating cancer, predicting prognosis, or determining the
 CC effectiveness of a chemotherapeutic or other treatment regimen. The
 CC expression vector may be used for the insertion and expression of CEA(6D)
 CC -1,2 nucleic acid encoding tumour antigens for the immunotherapeutic
 CC treatment of cancer. The target polypeptides are useful in generating
 CC antibodies used in screening assays or for immunotherapy. The present
 CC sequence represents the CEA(6D)-1,2 nucleotide sequence, which is given
 CC in the exemplification of the present invention.

XX
 SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-3 (1-9) x ADE13861 (1-2106)

QY 1 TyrluSerGlyAlaAspIleAsnLeu 9
 |||||
 Db 1810 TACCTTCGGGAGCGGACCTCAACCTC 1836

RESULT 9
 ADZ58977
 ID ADZ58977 standard; DNA; 2106 BP.
 XX
 AC ADZ58977;
 XX
 DT 30-JUN-2005 (first entry)
 DE Novel CEA(6)-1,2 expression vector-related mCEA DNA sequence SeqID7.
 XX
 DE expression; vector; CEA(6)-1,2; cytostatic; gene therapy; cancer; ds.
 XX
 OS Unidentified.
 XX
 PN WO2005035773-A2.
 XX
 PD 21-APR-2005.
 XX
 PF 06-OCT-2004; 2004WO-US033145.
 XX
 PR 08-OCT-2003; 2003US-0509593P.
 PR
 XX (AVET) AVENTIS PASTEUR INC.
 PA (THER-) THERION BIOLOGICS INC.
 XX
 XX Parrington M, Zhang L, Rovinski B, Gritz L, Greenhalgh P;
 XX WPI; 2005-296285/30.

XX New expression vector comprising the nucleic acid sequence CEA(6)-1,2 or
 PT its fragment, useful for preventing or treating cancer.

XX Example 1; SEQ ID NO 7; 72pp; English.

XX This invention relates to a novel expression vector containing the
 CC nucleic acid sequence CEA(6)-1,2. The invention may be useful for the
 CC development of compounds with a cytostatic activity whilst the disclosed
 CC sequences may be useful for gene therapy. The expression vector is useful
 CC for preventing or treating cancer. The present sequence is that of an
 CC mCEA expression vector DNA sequence which is related to the invention.
 CC Note: Another sequence was labelled as SeqID7 in the examples of the
 CC specification.

XX SQ Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-3 (1-9) x ADZ58977 (1-2106)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9

DB 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 10
 ADZ58978
 ID ADZ58978 standard; DNA; 2106 BP.
 XX AC ADZ58978;
 XX 30-JUN-2005 (first entry)
 DT Novel CEA(6)-1,2 expression vector-related mCEA DNA sequence SeqID8.
 DE expression; vector; CEA(6)-1,2; cytostatic; gene therapy; cancer; ds.
 KW Unidentified.
 XX OS
 XX PN WO2005035773-A2.
 XX PD 21-APR-2005.
 XX PF 06-OCT-2004; 2004WO-US033145.
 XX PR 08-OCT-2003; 2003US-0509593P.
 XX PA (AVET) AVENTIS PASTEUR INC.
 XX PA (THER-) THERION BIOLOGICS INC.
 XX PI Parrington M, Zhang L, Rovinski B, Gritz L, Greenhalgh P;
 XX WPI; 2005-296285/30.
 XX PT New expression vector comprising the nucleic acid sequence CEA(6)-1,2 or
 its fragment, useful for preventing or treating cancer.
 XX PS Example 1; SEQ ID NO 8; 72pp; English.

XX CC This invention relates to a novel expression vector containing the
 CC nucleic acid sequence CEA(6)-1,2. The invention may be useful for the
 CC development of compounds with a cytostatic activity whilst the disclosed
 CC sequences may be useful for gene therapy. The expression vector is useful
 CC for preventing or treating cancer. The present sequence is that of an
 CC mCEA expression vector DNA sequence which is related to the invention.
 CC Note: Another sequence was labelled as SeqID8 in the examples of the
 CC specification.
 XX SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-3 (1-9) x ADZ58978 (1-2106)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9

DB 1810 TACCTTCGGAGCGGACCTCAACCTC 1836

RESULT 11
 AEA11047
 ID AEA11047 standard; DNA; 2106 BP.
 XX AC AEA11047;
 XX 28-JUL-2005 (first entry)
 DT DNA encoding wobbled-CEA (wCEA).
 DE pancreas tumor; cytostatic; gastrointestinal disease; neoplasm; cancer;
 KW cytostatic; immune stimulation; gene therapy; immunotherapy; ds; gene;
 KW CEA; wCEA; carcinoembryonic antigen.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX FH Key
 XX CDS Location/Qualifiers
 988..2106
 /*tag= a
 /product= "wCEA"
 /partial
 /note= "No start codon"
 /transl_except= (pos:1825..1827,aa:Asn)

XX WO2005046622-A2.
 XX 26-MAY-2005.
 XX 12-NOV-2004; 2004WO-US038643.
 XX 12-NOV-2003; 2003US-0519354P.
 XX (THER-) THERION BIOLOGICS CORP.
 XX Panicali DL, Mazzara GP, Gritz LR;
 XX WPI; 2005-366921/37.
 XX P-PSDB; AEA11048.

XX PT Inducing an immunological response against a malignant pancreatic cell in
 an individual, useful for treating pancreatic cancer, by administering
 PT vectors containing genes that encode a pancreatic tumor-associated
 PT antigen (PTAA).
 XX PS Example 4; SEQ ID NO 3; 91pp; English.
 XX CC The invention relates to a method of inducing an immunological response
 CC against a malignant pancreatic cell in an individual, which comprises:
 CC selecting an individual having malignant pancreatic cells or at risk for
 CC developing such a pancreatic tumor, administering to the individual a
 CC first vector containing a first gene, or its antigenic portion, that
 CC encodes a pancreatic tumor-associated antigen (PTAA), and at regular
 CC intervals thereafter administering at least a second vector containing a
 CC gene encoding PTAA or its antigenic portion, where if carcinoembryonic
 CC antigen (CEA) or mucin 1 (MUC-1) or its antigenic portion or modified
 CC version is the PTAA, there must be a second PTAA present. The method
 CC further comprises administering granulocyte-macrophage colony stimulating
 CC factor (GM-CSF) or at least one co-stimulatory molecule. The co-
 CC stimulatory molecule is administered as a gene contained within the same
 CC or differing vector as the vector containing gene encoding the PTAA.
 CC PTAA, or its antigenic portion is contained in a poxvirus vector, such as
 CC vaccinia. The PTAA may be a mucin selected from MUC-1, MUC-2, MUC-3, MUC-
 CC 4, MUC-5AC, MUC-5B, MUC-6, MUC-7, MUC-11, MUC-12, and their antigenic
 CC portions and modified versions. The modified version is wobbled-MUC-1.
 CC The method is useful for treating individuals at risk of developing or
 CC suffering from pancreatic cancer. The present sequence represents DNA
 CC encoding wobbled-CEA (wCEA).
 XX SQ Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-3 (1-9) x AEA11047 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
 |||||
 Db 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 12

AEA12740
 ID AEA12740 standard; DNA; 2106 BP.

AC AEA12740;

DT 28-JUL-2005 (first entry)

DE Wobbled CEA DNA.

KW immunogenicity; breast tumor; Cytostatic; neoplasm; ds; gene; CEA.

OS Unidentified.

FH Key Location/Qualifiers

FT CDS 988..2106

FT /*tag= a

FT /product= "wCEA(6D)"

FT /transl_except= (pos:1825..1827,aa:Asn)

XX WO2005046614-A2.

PN 26-MAY-2005..

PD 12-NOV-2004; 2004WO-US037810.

PF 12-NOV-2003; 2003US-0519427P.

PR (THER-) THERION BIOLOGICS CORP.

PA Panicali DL, Mazzara GP, Gritz LR;

PI P-PSDB; AEA12741.

XX WPI; 2005-386205/39.

DR Inducing an immunological response against a cell expressing a breast cancer associated antigen in a human, useful for treating breast cancer by administering vectors containing genes that encode a breast cancer associated antigen.

XX Example 4; SEQ ID NO 3; 99pp; English.

XX The invention relates to a method of inducing an immunological response against a cell expressing a breast cancer associated antigen in a human comprises administering vectors containing genes or its antigenic portion that encode a breast cancer associated antigen. The method is useful for treating individuals suffering from breast cancer. The present sequence CC represents wobbled CEA DNA.

XX Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-3 (1-9) x AEA12740 (1-2106)

QY 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
 |||||
 Db 1810 TACCTTTCGGAGCGGACCTCAACCTC 1836

RESULT 13

AEB30768
 ID AEB30768 standard; DNA; 2106 BP.

XX AC AEB30768;

XX 06-OCT-2005 (first entry)

XX Multi-antigen construct CEA-CAP1-6D-1,2.

XX cytostatic; immunostimulant; vaccine; expression; vector; immunotherapy;

XX cancer; neoplasm; cellular transport; ds.

OS Unidentified.

PN WO2005068640-A2.

XX 28-JUL-2005.

XX 23-DEC-2004; 2004WO-US042980.

XX 23-DEC-2003; 2003US-0532205P.

XX (AVET) AVENTIS PASTEUR INC.

PA (THER-) THERION BIOLOGICS INC.

PI Parrington M, Berinstein N, Tartaglia JT, Panicali D, Gritz L;

XX WPI; 2005-522835/53.

XX Novel expression vector comprising nucleic acid sequences encoding modified KSA, useful for immunizing host, and for preventing or treating cancer.

XX Example 2; Fig 2; 64pp; English.

XX The invention describes an expression vector (I) useful for immunizing a host, comprising nucleic acid sequences encoding modified KSA. Also described are: a composition (CI) comprising (I) in a carrier, where (I) comprises nucleic acid sequences encoding modified KSA; an isolated DNA molecule comprising the modified KSA coding sequence of a fully defined approximately 945 nucleotide sequence (S1) given in the specification, and encoding modified KSA having a fully defined approximately 314 amino acid sequence given in the specification; and an isolated DNA molecule comprising carcinoembryonic antigen (CEA), p53, and modified KSA coding sequences, where the CEA sequence is CEA-CAP1-6D-1,2 having a fully defined approximately 2106 nucleotide sequence given in the specification, and the modified KSA sequence has (S1). A kit comprising (I) is also disclosed. (I) is useful for immunizing a host. (I) is useful for preventing or treating cancer, which involves administering (I) to a host, where (I) comprises nucleic acid sequences encoding modified KSA. This sequence represents CEA-CAP1-6D-1,2, a modified version of AEB30767 CC used in multi-antigen construct vcp2086.

XX Sequence 2106 BP; 574 A; 615 C; 435 G; 482 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 15 Length: 2106
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-3 (1-9) x AEB30768 (1-2106)

```
QY      1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db      1810 TACCTTTGGGAGCGGACCTCAACCTC 1836

RESULT 14
AEB30767
XX AEB30767 standard; DNA; 2106 BP.
XX AEB30767;
XX
XX 06-OCT-2005 (first entry)
XX
XX Multi-antigen construct CEA-CAP1-6D.
XX
XX cytostatic; immunostimulant; vaccine; expression; vector; immunotherapy;
XX cancer; neoplasm; cellular transport; ds.
XX
XX Unidentified.
XX
XX WO2005068640-A2.
XX
XX 28-JUL-2005.
XX
XX 23-DEC-2004; 2004WO-US042980.
XX
XX 23-DEC-2003; 2003US-0532205P.
XX
XX (AVET ) AVENTIS PASTEUR INC.
XX (THER-) THERION BIOLOGICS INC.
XX
XX Parrington M, Berinstein N, Tartaglia JT, Panicali D, Gritz L;
XX WPI; 2005-522835/53.
XX
XX Novel expression vector comprising nucleic acid sequences encoding
XX modified KSA, useful for immunizing host, and for preventing or treating
XX cancer.
XX
XX Example 2; Fig 2; 64pp; English.
XX
XX The invention describes an expression vector (I) useful for immunizing a
XX host, comprising nucleic acid sequences encoding modified KSA. Also
XX described are: a composition (CI) comprising (I) in a carrier, where (I)
XX comprises nucleic acid sequences encoding modified KSA; an isolated DNA
XX molecule comprising the modified KSA coding sequence of a fully defined
XX approximately 945 nucleotide sequence (S1) given in the specification,
XX and encoding modified KSA having a fully defined approximately 314 amino
XX acid sequence given in the specification; and an isolated DNA molecule
XX comprising carcinoembryonic antigen (CEA), p53, and modified KSA coding
XX sequences, where the CEA sequence is CEA-CAP1-6D-1,2 having a fully
XX defined approximately 2106 nucleotide sequence given in the
XX specification, the p53 sequence has a sequence as given in the
XX specification, and the modified KSA sequence has (S1). A kit comprising
XX CI is also disclosed. (I) is useful for immunizing a host. (I) is useful
XX for preventing or treating cancer, which involves administering (I) to a
XX host, where (I) comprises nucleic acid sequences encoding modified KSA.
XX This sequence represents CEA-CAP1-6D which is modified as shown in
XX AEB30768 for inclusion in multi-antigen construct vcp2086.
XX
XX Sequence 2106 BP; 559 A; 659 C; 442 G; 446 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.:      15      Length:      2106
Score:          43.00    Matches:      8
Percent Similarity: 100.00%  Conservative: 1
Best Local Similarity: 88.89%  Mismatches: 0
Query Match:      95.56%    Indels:      0
DB:              14      Gaps:      0

US-10-725-373-3 (1-9) x AEB30767 (1-2106)

QY      1 TyrLeuSerGlyAlaAspIleAsnLeu 9
```

```
Db      1810 TACCTTTGGGAGCGGACCTCAACCTC 1836

RESULT 15
AAI72490/C
ID AAI72490 standard; DNA; 7958 BP.
XX
XX AAI72490;
XX
XX 16-MAY-2002 (first entry)
XX
XX H6-promoter human CEAmod/42K-promoted B7.1 insertion cassette.
XX
XX CAP-1; epitope; carcinoembryonic antigen; CEA; agonist; immune response;
XX carcinoma; gastrointestinal; breast; pancreatic; bladder; ovarian; lung;
XX prostate; cancer; gene; ds.
XX
XX Homo sapiens.
XX Synthetic.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX misc_feature 423..827
XX /tag= a
XX /note= "ALVAC's C5 locus left flanking arm"
XX CDS complement(903..3008)
XX /tag= b
XX /product= "CEA agonist peptide"
XX promoter complement(3009..3132)
XX /tag= c
XX /label= Vaccinia_H6_promoter
XX promoter 3210..3275
XX /tag= d
XX /label= 42K_promoter
XX CDS 3276..4142
XX /tag= e
XX /product= "Human B7.1"
XX misc_feature 4184..5722
XX /tag= f
XX /note= "ALVAC's C5 locus right flanking arm#"
XX
XX WO200210379-A2.
XX
XX 07-FEB-2002.
XX
XX 27-JUL-2001; 2001WO-CA001092.
XX
XX 31-JUL-2000; 2000US-0222043P.
XX (AVET ) AVENTIS PASTEUR LTD.
XX (THER-) THERION BIOLOGICS.
XX (USSH ) US NAT CANCER INST.
XX
XX Berinstein N, Tartaglia J, Tine JA, Panicali DL, Gritz L;
XX Schlom J;
XX WPI; 2002-206189/26.
XX
XX Carcinoembryonic antigen agonist polypeptide for inducing an immune
XX response in animal against antigen and for inhibiting an epitope antigen
XX expressing carcinoma cell, comprises a modified antigen epitope.
XX
XX Example 3; Fig 4; 69pp; English.
XX
XX This sequence represents ALVAC(2)-CEAmod/hb7.1. This is a coding sequence
XX containing the H6 promoted modified carcinoembryonic antigen (CEA)
XX agonist polypeptide of the invention. The CEA agonist contains a modified
XX CAP-1 epitope of CEA, in which position 6 is modified from Asp to Asn to
XX increase its immunogenicity. The CEA agonist polypeptide of the
XX invention, or DNA encoding it, are useful for: (i) inducing an immune
XX response in an animal directed against a CEA protein or fragment, CEA
XX agonist, a CEA epitope, a modified CEA epitope, cells expressing or
XX binding a CEA protein or fragment; and (ii) inhibiting a CEA epitope
```

CC expressing carcinoma cell, which is a gastrointestinal, breast,
CC pancreatic, bladder, ovarian, lung or prostate carcinoma cell in a
CC patient, hence is useful for manufacture of a medicament for the
CC treatment of cancer

XX
SQ Sequence 7958 BP; 2096 A; 1720 C; 1858 G; 2284 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 67.9 Length: 7958
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-3 (1-9) x AAI72490 (1-7958)

Qy 1 TyrLeuSerGlyAlaAspIleAsnLeu 9
Db 1199 TACCTTTCGGGAGCGGACCTCAACCTC 1173

Search completed: December 6, 2005, 16:29:53
Job time : 371.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 10:17:58 ; Search time 367.25 Seconds
(without alignments)
163.328 Million cell updates/sec

Title: US-10-725-373-4
Perfect score: 45
Sequence: 1 YLSGANINL 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh
-Q=/Cgn2_1/USPTO_spool/US1075373/runat_01122005_114444_21420/app_query.fasta_1.796
-DB=N Geneseq -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1075373 @CGN_1_1_1244 @runat_01122005_114444_21420 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	95.6	27	2 AAX56260	Aax56260 Carcinoem
2	43	95.6	30	12 ADL46174	Adl46174 Human CAP
3	43	95.6	64	12 ADL46175	Adl46175 Human imm
4	43	95.6	80	2 AAV57948	AAV57948 708 vkcea

C	5	43	95.6	80	2 AAV81101	Aav81101 Vaccine 2
	6	43	95.6	155	4 Aai29234	Aai29234 Colon tum
	7	43	95.6	155	8 ABZ33420	Abz33420 Human col
	8	43	95.6	256	4 Aas57750	Aas57750 cDNA #426
C	9	43	95.6	340	6 ABV88334	Abv88334 Human col
	10	43	95.6	402	14 ACL62053	ACL62053 Human col
C	11	43	95.6	407	4 Aas57366	Aas57366 cDNA #42
C	12	43	95.6	409	4 AAS57425	Aas57425 cDNA #101
	13	43	95.6	409	6 ABV86774	ABV86774 Human col
C	14	43	95.6	409	6 ABV87551	ABV87551 Human col
	15	43	95.6	409	6 ABV89100	ABV89100 Human col
	16	43	95.6	409	6 ABV87855	ABV87855 Human col
	17	43	95.6	409	6 ABK39290	ABK39290 DNA encod
	18	43	95.6	409	6 ABK39002	ABK39002 cDNA enco
C	19	43	95.6	409	6 ABK39424	ABK39424 DNA enco
	20	43	95.6	409	6 ABK45946	ABK45946 cDNA enco
	21	43	95.6	409	6 ABK45289	ABK45289 cDNA enco
C	22	43	95.6	409	6 ABK27782	ABk27782 Human col
	23	43	95.6	409	8 ACA11619	ACA11619 Human lun
	24	43	95.6	409	8 ACA11331	ACA11331 Human lun
C	25	43	95.6	409	8 ACA11753	ACA11753 Human lun
	26	43	95.6	409	8 ACA02939	ACA02939 Lung canc
C	27	43	95.6	409	8 ACA02805	ACA02805 Lung canc
	28	43	95.6	409	8 ACA02517	ACA02517 Lung canc
	29	43	95.6	409	10 ADH46559	Adh46559 Human lun
	30	43	95.6	409	10 ADH46847	Adh46847 Human lun
C	31	43	95.6	409	10 ADH46981	Adh46981 Human lun
	32	43	95.6	409	13 ADJ20766	Adj20766 Human lun
	33	43	95.6	409	13 ADJ20478	Adj20478 Human lun
C	34	43	95.6	409	13 ADJ20900	Adj20900 Human lun
C	35	43	95.6	410	4 AAS57430	Aas57430 cDNA #106
	36	43	95.6	410	4 AAS57434	Aas57434 cDNA #110
C	37	43	95.6	410	6 ABV88264	ABv88264 Human col
	38	43	95.6	412	3 AAA77726	Aaa77726 cDNA enco
	39	43	95.6	412	4 AAI28464	Aai28464 Colon tum
	40	43	95.6	412	8 ABZ32650	ABz32650 Human col
	41	43	95.6	413	4 AAS57627	Aas57627 cDNA #303
	42	43	95.6	415	5 AAF68500	Aaf68500 Human lun
	43	43	95.6	415	6 ABK38411	ABk38411 cDNA enco
	44	43	95.6	415	8 ACA10740	ACA10740 Human lun
	45	43	95.6	415	8 ABX99691	ABx99691 Lung canc

ALIGNMENTS

RESULT 1
AAX56260
ID AAX56260 standard; DNA; 27 BP.
XX
AC AAX56260;
XX
DT 20-JUL-1999 (first entry)
XX
DE Carcinoembryonic antigen peptide agonist CAP-1 encoding DNA SEQ ID NO:6.
XX
KW Carcinoembryonic antigen; CEA; human; agonist; antagonist;
XX immune response; carcinoma; gastrointestinal; breast; pancreatic;
XX bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
XX adoptive transfer therapy; autoimmune reaction; immunotherapy; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9919478-A1.
XX
PD 22-APR-1999.
XX
PF 22-SEP-1998; 98WO-US019794.
XX
PR 10-OCT-1997; 97US-0061589P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX

PI Schlom J, Barzaga E, Zaremba S;
 XX WPI; 1999-326544/27.
 XX Peptide agonists and antagonists of carcinoembryonal antigen.
 XX Disclosure; Page 19; 72pp; English.
 XX The present invention describes peptides (A) that comprise agonists (Ia)
 CC or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
 CC used in vaccines to kill or inhibit carcinoma cells that express CEA or
 CC its epitopes, particularly for treating gastrointestinal, breast,
 CC pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
 CC be used to proliferate T cells, e.g. from vaccinated subjects, for use in
 CC adoptive transfer therapy. (Ib) are used to inhibit CEA-specific immune
 CC responses, e.g. in vaccinated subjects, to prevent an autoimmune reaction
 CC to cancer immunotherapy (i.e. to prevent attack on normal but CEA-
 CC expressing cells). (Ia) are more active than native sequence (I) and
 CC generate a highly specific and systemic anti-CEA response. Cytotoxic T
 CC cells generated recognize both (Ia) and native CEA epitopes. The present
 CC sequence encodes a specifically claimed example of (Ia)
 XX
 SQ Sequence 27 BP; 6 A; 10 C; 5 G; 6 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.106 Length: 27
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 2 Gaps: 0
 US-10-725-373-4 (1-9) x AAX56260 (1-27)
 QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
 DB 1 TACCTTTTCGGGAGCGAACCTCAACCTC 27
 RESULT 2
 ADL46174
 ID ADL46174 standard; DNA; 30 BP.
 AC ADL46174;
 XX
 DT 17-JUN-2004 (first entry)
 XX
 DE Human CAP-1 tumour antigen fragment DNA, SEQ ID NO:8 #1.
 XX
 KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
 KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
 KW human; CAP-1; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2004024181-A1.
 XX
 PD 25-MAR-2004.
 XX
 PF 15-SEP-2003; 2003WO-CN000776.
 XX
 PR 13-SEP-2002; 2002CN-00136965.
 XX
 PA (LIJJ/) LI J.
 XX
 PI Li J;
 XX
 DR WPI; 2004-269898/25.
 XX
 XX Tumor-antigen vaccines with molecular weight far smaller than antigen-
 PT antibody compound to enable endocytosis by dendritic cells to promote
 XX very high immunoreaction for killing tumor cells.
 XX
 PS Example 2; SEQ ID NO 8; 28pp; Chinese.

XX The invention relates to a tumour antigen vaccine comprising 7 or more
 CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
 CC fragment. The invention also relates to DNA sequences encoding the
 CC antigenic fusion polypeptide; expression vectors and host cells
 CC comprising the DNA sequences; a process for recombinantly producing the
 CC fusion polypeptide; and a vaccine composition comprising the fusion
 CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
 CC used is preferably selected from 07-AP, APP, ART-4, BAGE B, beta-
 CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, CYP-B,
 CC DAM, ELF2M, ETv6-AML1, ETS, G250, GAGE, Gnt-V, GP100, HAGE, HER-2/NEU,
 CC HLA-A*0201-R1701, HPV-E6, EBNA, HSP70-2M, HST-2, hTERT, ICS,
 CC KIAA0205, LAGE, LDLR/FUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
 CC Myosin/m, MUC1, MUM-1-2,-3, NA88-A, NY-ESO-1, P15, p190, P53, Pml/RAR
 CC alpha, FRAME, PSA, PSM, RAGE, RAS, RUL, RUL2, SAGE, SART-1, SART-3,
 CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
 CC antigen vaccines of the invention are useful in cancer therapy. The
 CC antigenic fusion protein used in the vaccine are much smaller than the
 CC corresponding antibody-antigen complex, permitting them to be endocytosed
 CC by dendritic cells and thereby resulting in a greatly increased anti-
 CC tumour immune response. The present sequence represents DNA encoding a
 CC fragment of the human CAP-1/CEA tumour antigen used in an example of the
 CC invention. Note: The present sequence differs from that also referred to
 CC as SEQ ID NO:8 () which is given on page 10 of the specification.
 XX
 SQ Sequence 30 BP; 6 A; 12 C; 5 G; 7 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.12 Length: 30
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 12 Gaps: 0
 US-10-725-373-4 (1-9) x ADL46174 (1-30)
 QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
 DB 1 TACCTTTTCGGGAGCGAACCTCAACCTC 27
 RESULT 3
 ADL46175
 ID ADL46175 standard; DNA; 64 BP.
 AC ADL46175;
 XX
 DT 17-JUN-2004 (first entry)
 XX
 DE Human immunoglobulin Fc fragment 5' PCR primer, SEQ ID NO:9 #1.
 XX
 KW Tumour antigen; vaccine; immunoglobulin; CH3 fragment; fusion protein;
 KW cancer; tumour; dendritic cell; endocytosis; immune response; cytostatic;
 KW human; Fc fragment; PCR; primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2004024181-A1.
 XX
 PD 25-MAR-2004.
 XX
 PF 15-SEP-2003; 2003WO-CN000776.
 XX
 PR 13-SEP-2002; 2002CN-00136965.
 XX
 PA (LIJJ/) LI J.
 XX
 PI Li J;
 XX
 DR WPI; 2004-269898/25.
 XX
 XX Tumor-antigen vaccines with molecular weight far smaller than antigen-
 PT antibody compound to enable endocytosis by dendritic cells to promote

PT very high immunoreaction for killing tumor cells.

XX Example 2; SEQ ID NO 9; 28bp; Chinese.

XX The invention relates to a tumour antigen vaccine comprising 7 or more
 CC amino acids of a tumour antigen sequence joined to an immunoglobulin CH3
 CC fragment. The invention also relates to DNA sequences encoding the
 CC antigenic fusion polypeptide; expression vectors and host cells
 CC comprising the DNA sequences; a process for recombinantly producing the
 CC fusion polypeptide; and a vaccine composition comprising the fusion
 CC polypeptide and a pharmaceutically acceptable carrier. The tumour antigen
 CC used is preferably selected from 07-AB, AFP, ART-4, BAGE B, beta-
 CC catenin/m, bcr-abl, CAMEL, CAP-1, CASP-8, CDC27m, CDK4/m, CEA, CT, Cyp-B,
 CC DAM, ELF2M, ETV6-AML1, ETS, G250, GAGE, GNT-V, GP100, HAGE, HER-2/NEU,
 CC HLA-A*0201-R1701, HPV-E6, HPV-E7, EBNA, HSP70-2M, HST-2, HTERT, ICE,
 CC KIAA0205, LAGE, LDR/LPUT, GDP-Lfucose, MAGE, MART-1/Melan-A, MCIR,
 CC Myosin/m, MUC1, MUM-1, -2, -3, NAB8-A, NY-ESO-1, P15, P190, P53, Pml/RAR
 CC alpha, FRAME, PSA, PSM, RAS, RAS, RUI, RU2, SAGE, SART-1, SART-3,
 CC TEL/AML1, TPI/m, TRP-1, gp75, TRP-2, TRP-2/INT2 and WTI. The tumour
 CC antigen vaccines of the invention are useful in cancer therapy. The
 CC corresponding antibody-antigen complex, permitting them to be endocytosed
 CC by dendritic cells and thereby resulting in a greatly increased anti-
 CC tumour immune response. Sequences ADL46175-ADL46176 represent PCR primers
 CC used to amplify DNA encoding a human immunoglobulin Fc fragment in an
 CC example of the invention. Note: The present sequence differs from that
 CC also referred to as SEQ ID NO:9 () which is given on page 10 of the
 CC specification.

XX Sequence 64 BP; 16 A; 22 C; 13 G; 13 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.293 Length: 64
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 12 Gaps: 0

US-10-725-373-4 (1-9) x ADL46175 (1-64)

QY 1 TyrLeuSerGlyAlaAseNleAseNLeu 9
 Db 15 TACCTTCGGAGCGACCTCACTTC 41

RESULT 4

AAV57948/C
 ID AAV57948 standard; DNA; 80 BP.

AC AAV57948;

XX 24-NOV-1998 (first entry)

XX 708 vkcea primary reaction 1.2 oligonucleotide vkcea592r.

XX Hepatitis B surface antigen; HBsAg; MHC class II-restricted peptide;
 KW vaccination; vaccine; MHC class I molecule; immune response; cancer;
 KW major histocompatibility complex molecule; pathogenic organism;
 KW viral disease; autoimmune condition; allergy; PCR primer; ss.

XX Synthetic.

XX WO9833523-A1.

XX 06-AUG-1998.

XX 02-FEB-1998; 98WO-GB000325.

XX 31-JAN-1997; 97GB-00001999.

PR 05-JUL-1997; 97GB-00014182.

PR 07-AUG-1997; 97GB-00016620.

PR 07-AUG-1997; 97GB-00016641.

PR 21-NOV-1997; 97GB-00024584.

XX (BIOV-) BIOVATION LTD.

XX Carr FJ, Carter G;

XX WPI; 1998-437178/37.

XX Immunogenic molecules - comprising nucleic acid and polypeptide portion,
 PT from both of which peptide for presentation on major histocompatibility
 PT complex molecules can be derived.

XX Example 10; Page 60; 87pp; English.

XX A molecule has been developed which comprises: (a) a nucleic acid portion
 CC from which at least one peptide for presentation of MHC class I or class
 CC II molecules, or both, may be derived, and (b) a polypeptide portion,
 CC from which at least 1 peptide for presentation on MHC class I or class II
 CC molecules, or both, may be derived. Also described in the present
 CC invention is another molecule comprising: (a) a nucleic acid portion from
 CC which at least 1 peptide for presentation on MHC class I or class II
 CC molecules, or both, may be derived, and (b) a polypeptide portion
 CC comprising a recognition domain capable of targeting the molecule to an
 CC antigen presenting cell (APC), where the polypeptide portion does not
 CC comprise a specific antigen binding site. The molecules can be used to
 CC induce immune responses to treat or prevent, e.g. diseases caused by
 CC pathogenic organisms, cancers, viral disease, e.g. HIV or hepatitis
 CC infection, autoimmune conditions, e.g. Grave's disease, multiple
 CC sclerosis, systemic lupus erythematosus, diabetes mellitus, Kawasaki's
 CC disease, rheumatoid arthritis or allergies, e.g. atopic dermatitis,
 CC allergic rhinitis, allergic conjunctivitis, atopic asthma or eczema. The
 CC combination of DNA and polypeptide in the same molecule can give rise not
 CC only to a combination of MHC class I- and MHC class II-mediated immune
 CC responses but also to an enhancement of these responses compared to the
 CC responses given by either DNA or polypeptide alone. The present sequence
 CC represents an oligonucleotide used in an example from the present
 CC invention

XX Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.381 Length: 80
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 2 Gaps: 0

US-10-725-373-4 (1-9) x AAV57948 (1-80)

QY 1 TyrLeuSerGlyAlaAseNleAseNLeu 9

Db 68 TACCTGTCCGGCGCAACCTGACCTG 42

RESULT 5

AAV81101/C

ID AAV81101 standard; DNA; 80 BP.

XX AAV81101;

XX 03-MAR-1999 (first entry)

XX Vaccine 2 708 VI constructing long oligo VKCEA592R.

XX Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
 KW immunoglobulin; therapeutic; streptokinase; vaccine; 708; ss.

XX Synthetic.

XX WO9852976-A1.

XX 26-NOV-1998.

XX 21-MAY-1998; 98WO-GB001473.

XX 21-MAY-1997; 97GB-00010480.
 PR 31-JUL-1997; 97GB-00016197.
 PR 28-NOV-1997; 97GB-00025270.
 PR 02-DEC-1997; 97US-0067235P.
 PR 14-APR-1998; 98GB-00007751.
 XX (BIOV-) BIOVATION LTD.
 PA Carr FU;
 PI WPI; 1999-045301/04.
 XX Reducing immunogenicity of proteins - by modifying the amino acid
 PT sequence of the protein to eliminate potential epitopes for T-cells of a
 PT given species.
 XX Example 4; Fig 20; 77pp; English.
 PS The invention relates to a method for the production of non-immunogenic
 CC proteins. The method comprises determining at least part of the amino
 CC acid sequence of the protein; (b) identifying in the amino acid sequence
 CC one or more potential epitopes for T-cells (T-cell epitopes) of the given
 CC species; and (c) modifying the amino acid sequence to eliminate at least
 CC one of the T-cell epitopes identified in step (b) thereby to eliminate or
 CC reduce the immunogenicity of the protein when exposed to the immune
 CC system of the given species. A method of analysing a pre-existing protein
 CC to predict the basis for immunogenic responses is also provided. The
 CC methods can be used particularly for reducing the immunogenicity of
 CC immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
 CC products can be used for diagnosis and therapy. Sequences AAV81090-110
 CC represent oligonucleotides used for the construction of vaccine 2 708 Vh
 CC and V1
 XX Sequence 80 BP; 16 A; 28 C; 19 G; 17 T; 0 U; 0 Other;
 SQ Alignment Scores:
 Pred. No.: 0.381 Length: 80
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 2 Gaps: 0
 US-10-725-373-4 (1-9) x AAV81101 (1-80)
 Qy 1 TyrLeuSerGlyAlaAsnIleAenLeu 9
 Db 68 TACCTGTCCGGCGCCCAACTGAACTG 42
 RESULT 6
 AAI29234
 ID AAI29234 standard; cDNA; 155 BP.
 AC AAI29234;
 XX 12-OCT-2001 (first entry)
 DT Colon tumour related determined cDNA sequence for clone R0094:D08.
 DE Human; immunotherapy; diagnosis; colon cancer; colon tumour; immunogenic;
 KW gene therapy; vaccine; colonic cancer; ss.
 KW Homo sapiens.
 OS WO200149716-A2.
 XX 12-JUL-2001.
 PD 29-DEC-2000; 2000WO-US035596.
 PF 30-DEC-1999; 99US-00476296.
 PR 10-JAN-2000; 2000US-00480321.
 PR

PR 15-FEB-2000; 2000US-00504629.
 PR 06-MAR-2000; 2000US-00519444.
 PR 19-MAY-2000; 2000US-00575251.
 PR 29-JUN-2000; 2000US-00609448.
 PR 28-AUG-2000; 2000US-00649811.
 XX (CORI-) CORIXA CORP.
 PA Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 XX WPI; 2001-441847/47.
 DR Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer.
 XX Claim 2; Page 356; 472pp; English.
 PS The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be used
 CC to treat disorders associated with decreased expression by rectifying
 CC mutations or deletions in a patient's genome that affect the activity of
 CC TCAPs by expressing inactive proteins or to supplement the patient's own
 CC production of them. Additionally, (II) may be used to produce the TCAP
 CC proteins, by inserting the nucleic acids into a host cell culturing the
 CC cell to express the protein. (II) and its complementary sequences may
 CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
 CC and hybridisation assays to detect and quantitate the presence of similar
 CC nucleic acids in samples, and therefore which patients may be in need of
 CC restorative therapy. (I) may also be used as antigens in the production
 CC of antibodies against TCAPs and in assays to identify modulators of TCAP
 CC expression and activity. Anti-(I) antibodies and antagonists may also be
 CC used to down regulate TCAP expression and activity. The anti-(I)
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). AAI28460 to AAI29512 and AAM24494 to AAM24523 represent
 CC nucleotide and amino acid sequences given in the exemplification of the
 CC present invention
 XX Sequence 155 BP; 30 A; 56 C; 31 G; 30 T; 0 U; 8 Other;
 SQ Alignment Scores:
 Pred. No.: 0.827 Length: 155
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 4 Gaps: 0
 US-10-725-373-4 (1-9) x AAI29234 (1-155)
 Qy 1 TyrLeuSerGlyAlaAsnIleAenLeu 9
 Db 119 TACCTTTCGAGCGAAGCACTCAACCTC 145
 RESULT 7
 AB233420
 ID AB233420 standard; cDNA; 155 BP.
 XX AC AB233420;
 XX 30-JAN-2003 (first entry)
 DT Human colon tumour cDNA for clone R0094:D08 SEQ ID NO:788.
 XX Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
 KW tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
 XX Homo sapiens.
 OS

XX WO200283070-A2.
XX
XX
XX PD 24-OCT-2002.
XX
XX PF 09-APR-2002; 2002WO-US011475.
XX
XX PR 10-APR-2001; 2001US-00833263.
XX PR 03-AUG-2001; 2001US-00922217.
XX PR 19-DEC-2001; 2001US-00025380.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD, Skeiky YAW;
PI Fanger GR, Vedvick TS, Carter D;
XX
XX DR WPI; 2003-067548/06.
XX
XX PT New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer.
XX
XX PS Disclosure; Page 357; 537pp; English.
XX
XX CC The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting the
CC presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (1) have immunostimulant and
CC cytostatic activities and can be used in vaccines. AB232646 to AB233725
CC and ABP55343 to ABP5391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention
XX
XX SQ Sequence 155 BP; 30 A; 56 C; 31 G; 30 T; 0 U; 8 Other;

Alignment Scores:
Pred. No.: 0.827 Length: 155
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 8 Gaps: 0

US-10-725-373-4 (1-9) x AB233420 (1-155)

QY 1 TyrLeuSerGlyAlaAsnLeu 9
DB 119 TACCTTTTCGGAGCGAACCTCAACCTC 145

RESULT 8
AAS57750
ID AAS57750 standard; cDNA; 256 BP.
XX
XX AC AAS57750;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE cDNA #426 encoding portion of a human colon tumour protein.
XX
XX KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200173027-A2.
XX
XX PD 04-OCT-2001.
XX
XX PF 22-MAR-2001; 2001WO-US009246.
XX
XX XX 24-MAR-2000; 2000US-0191597P.

PR 04-MAY-2000; 2000US-0202024P.
PR 05-MAY-2000; 2000US-0202189P.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Meagher MJ, Xu J, King GE;
XX
XX DR WPI; 2001-611627/70.
XX
XX PT New colon tumor proteins and related nucleic acid, useful for treatment,
PT prevention, diagnosis and monitoring of cancer.
XX
XX PS Claim 4; Page 125; 299pp; English.
XX
XX CC Th present invention relates to the isolation of novel cDNA sequences
CC encoding for at least an immunogenic portion of human colon tumour
CC proteins. The sequences of the invention are useful in pharmaceutical
CC compositions and vaccines for the prevention and treatment of cancers
CC such as colon cancer. They are also useful for the diagnosis and
CC monitoring of such cancers. Antibodies to the colon tumour proteins and
CC antigen presenting cells that express polynucleotides encoding colon
CC tumour proteins can be used to inhibit the development of cancers. T-
CC cells that react specifically with colon tumour proteins are useful for
CC removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC AAS5880 represent the cDNA sequences of the invention that encode for
CC portions of human colon tumour proteins
XX
XX SQ Sequence 256 BP; 62 A; 84 C; 44 G; 66 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.49 Length: 256
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 4 Gaps: 0

US-10-725-373-4 (1-9) x AAS57750 (1-256)

QY 1 TyrLeuSerGlyAlaAsnLeu 9
DB 15 TACCTTTTCGGAGCGAACCTCAACCTC 41

RESULT 9
ABV88334/c
ID ABV88334 standard; cDNA; 340 BP.
XX
XX AC ABV88334;
XX
XX DT 13-DEC-2002 (first entry)
XX
XX DE Human colon cancer related cDNA SEQ ID NO 1645.
XX
XX KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
XX ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200258534-A2.
XX
XX PD 01-AUG-2002.
XX
XX PF 16-NOV-2001; 2001WO-US043704.
XX
XX PR 20-NOV-2000; 2000US-0252222P.
XX PR 06-FEB-2001; 2001US-0267011P.
XX PR 28-MAR-2001; 2001US-0279670P.
XX PR 10-JUL-2001; 2001US-0304037P.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;

XX WPI; 2002-608400/65.
XX New isolated tumor colon polynucleotide and polypeptide, useful for the
PT diagnosis, prevention and/or treatment of cancer, in particular colon
PT cancer.
XX
XX Claim 1; SEQ ID NO 1645; 266pp + Sequence Listing; English.
XX
XX The invention relates to a human colon tumour expressed polynucleotide
CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC sequences that hybridize to (i), under moderately stringent conditions;
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC degenerate variants of (i). The compositions and methods of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (I) can be used in gene therapy and
CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 340 BP; 82 A; 63 C; 114 G; 81 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.08 Length: 340
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 6 Gaps: 0
US-10-725-373-4 (1-9) x ABV86334 (1-340)
QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
Db 241 TACCTTCGGAGCGAACCTCAACCTC 215
RESULT 10
ID ACL62053 standard; cDNA; 402 BP.
XX
XX ACL62053;
XX
XX 24-MAR-2005 (first entry)
XX Human colon cancer differentially expressed polynucleotide, SEQ ID:8188.
XX
XX Differential expression; diagnosis; therapy; drug screening; cancer;
KW neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;
KW ss.
XX Homo sapiens.
XX
XX WO2005000087-A2.
XX
XX 06-JAN-2005.
XX
XX 13-MAY-2004; 2004WO-US015421.
XX
XX 03-JUN-2003; 2003US-0475872P.
XX
XX (CHIR) CHIRON CORP.
XX Randazzo F, Moler E, Escobedo J, Garcia PD;
XX WPI; 2005-075421/08.
XX New isolated polynucleotides, which are differentially expressed in colon
PT cancer cell, useful for treating cancer, e.g. colon cancer, breast
PT cancer, or pancreatic cancer.
XX

PS Claim 1; SEQ ID NO 8188; 97pp; English.
XX
XX The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which
CC are differentially expressed in colon cancer cells. The invention also
CC relates to vectors and host cells comprising a differentially expressed
CC polynucleotide of the invention; a method for detecting a cancerous cell
CC by detection of a gene product of the polynucleotides; a method for
CC inhibiting a cancerous phenotype of a cell by inhibiting a gene product
CC of the polynucleotides; a method of treating an individual with cancer by
CC administration of a modulator of a gene product of the polynucleotides;
CC and an isolated antibody that specifically binds to a polypeptide encoded
CC by one of the 9672 polynucleotides. The polynucleotides, polypeptides,
CC antibodies, and methods are useful for the detection of cancerous cells;
CC for the diagnosis, prognosis and management of cancer; for the
CC identification of agents that modulate the phenotype of cancerous cells;
CC and for the treatment of cancer, especially colon cancer and metastasized
CC colon cancer, but also breast or pancreatic cancer. The polynucleotides
CC are also useful as a source of probes or primers for use in diagnostic
CC methods. The differentially expressed polynucleotides or their encoded
CC proteins can additionally be used as vaccines to modulate primary immune
CC responses for the prevention or treatment of cancer. The present sequence
CC represents a specifically claimed polynucleotide which is differentially
CC expressed in colon cancer. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 402 BP; 84 A; 122 C; 83 G; 113 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.53 Length: 402
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 14 Gaps: 0
US-10-725-373-4 (1-9) x ACL62053 (1-402)
QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
Db 57 TACCTTCGGAGCGAACCTCAACCTC 83
RESULT 11
AAS57366/C
ID AAS57366 standard; cDNA; 407 BP.
XX
XX AAS57366;
XX
XX 13-FEB-2002 (first entry)
XX
XX cDNA #42 encoding portion of a human colon tumour protein.
XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX Homo sapiens.
XX
XX WO200173027-A2.
XX
XX 04-OCT-2001.
XX
XX 22-MAR-2001; 2001WO-US009246.
XX
XX 24-MAR-2000; 2000US-0191597P.
PR 04-MAY-2000; 2000US-0202024P.
PR 05-MAY-2000; 2000US-0202189P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Meagher MJ, Xu J, King GE;
PI
XX WPI; 2001-611627/70.
DR

XX New colon tumor proteins and related nucleic acid, useful for treatment,
PT prevention, diagnosis and monitoring of cancer.
PS Claim 4; Page 68; 299pp; English.
XX Th present invention relates to the isolation of novel cDNA sequences
CC encoding for at least an immunogenic portion of human colon tumour
CC proteins. The sequences of the invention are useful in pharmaceutical
CC compositions and vaccines for the prevention and treatment of cancers
CC such as colon cancer. They are also useful for the diagnosis and
CC monitoring of such cancers. Antibodies to the colon tumour proteins and
CC antigen presenting cells that express polynucleotides encoding colon
CC tumour proteins can be used to inhibit the development of cancers. T-
CC cells that react specifically with colon tumour proteins are useful for
CC removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC AAS58880 represent the cDNA sequences of the invention that encode for
CC portions of human colon tumour proteins
XX SQ Sequence 407 BP; 97 A; 76 C; 130 G; 100 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 2.57 Length: 407
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 4 Gaps: 0

US-10-725-373-4 (1-9) x AAS57366 (1-407)

QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
DB 241 TACCTTTCGGAGCGAACCTCAACCTC 215

RESULT 12
AAS57425/c
ID AAS57425 standard; cDNA; 409 BP.
XX AC AAS57425;
XX DT 13-FEB-2002 (first entry)
XX DE cDNA #101 encoding portion of a human colon tumour protein.
XX KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200173027-A2.
XX PD 04-OCT-2001.
XX PF 22-MAR-2001; 2001WO-US009246.
XX PR 24-MAR-2000; 2000US-0191597P.
XX PR 04-MAY-2000; 2000US-0202024P.
XX PR 05-MAY-2000; 2000US-0202189P.
XX PA (CORI-) CORIXA CORP.
XX PI Meagher MJ, Xu J, King GE;
XX WPI; 2001-611627/70.
XX New colon tumor proteins and related nucleic acid, useful for treatment,
PT prevention, diagnosis and monitoring of cancer.
PS Claim 4; Page 77; 299pp; English.
XX Th present invention relates to the isolation of novel cDNA sequences
CC encoding for at least an immunogenic portion of human colon tumour

CC proteins. The sequences of the invention are useful in pharmaceutical
CC compositions and vaccines for the prevention and treatment of cancers
CC such as colon cancer. They are also useful for the diagnosis and
CC monitoring of such cancers. Antibodies to the colon tumour proteins and
CC antigen presenting cells that express polynucleotides encoding colon
CC tumour proteins can be used to inhibit the development of cancers. T-
CC cells that react specifically with colon tumour proteins are useful for
CC removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC The polynucleotides sequences are also useful in gene therapy. AAS57325-
CC AAS58880 represent the cDNA sequences of the invention that encode for
CC portions of human colon tumour proteins
XX SQ Sequence 409 BP; 98 A; 76 C; 132 G; 103 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.58 Length: 409
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.56% Indels: 0
DB: 4 Gaps: 0

US-10-725-373-4 (1-9) x AAS57425 (1-409)

QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
DB 241 TACCTTTCGGAGCGAACCTCAACCTC 215

RESULT 13
ABV86774
ID ABV86774 standard; cDNA; 409 BP.
XX AC ABV86774;
XX DT 13-DEC-2002 (first entry)
XX DE Human colon cancer related cDNA SEQ ID NO 85.
XX KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
XX OS Homo sapiens.
XX PN WO200258534-A2.
XX PD 01-AUG-2002.
XX PF 16-NOV-2001; 2001WO-US043704.
XX PR 20-NOV-2000; 2000US-0252222P.
XX PR 06-FEB-2001; 2001US-0267011P.
XX PR 28-MAR-2001; 2001US-0279670P.
XX PR 10-JUL-2001; 2001US-0304037P.
XX PA (CORI-) CORIXA CORP.
XX PI Scolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
XX WPI; 2002-608400/65.
XX New isolated tumor colon polynucleotide and polypeptide, useful for the
PT diagnosis, prevention and/or treatment of cancer, in particular colon
PT cancer.
XX Claim 1; SEQ ID NO 85; 266pp + Sequence Listing; English.
XX The invention relates to a human colon tumour expressed polynucleotide
CC (i) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC sequences that hybridize to (i), under moderately stringent conditions;
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC degenerate variants of (i). The compositions and methods of the present

CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 BP; 103 A; 132 C; 76 G; 98 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.58 Length: 409
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 6 Gaps: 0

US-10-725-373-4 (1-9) x ABV86774 (1-409)

QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
 DB 169 TACCTTTCGGAGCGAACCCTCAACCTC 195

RESULT 14
 ABV87551/c
 ID ABV87551 standard; cDNA; 409 BP.

XX AC ABV87551;
 XX DT 13-DEC-2002 (first entry)
 XX DE Human colon cancer related cDNA SEQ ID NO 862.
 XX KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 XX SS.
 XX OS Homo sapiens.
 XX PN WO200258534-A2.
 XX PD 01-AUG-2002.
 XX PF 16-NOV-2001; 2001WO-US043704.
 XX PR 20-NOV-2000; 2000US-0252222P.
 XX PR 06-FEB-2001; 2001US-0267011P.
 XX PR 28-MAR-2001; 2001US-0279670P.
 XX PR 10-JUL-2001; 2001US-0304037P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
 XX WPI; 2002-608400/65.

PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer.

PS Claim 1; SEQ ID NO 862; 266pp + Sequence Listing; English.

CC The invention relates to a human colon tumour expressed polynucleotide
 CC (i) encoding a polypeptide (ii), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 409 BP; 98 A; 76 C; 132 G; 103 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.58 Length: 409
 Score: 43.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 88.89% Mismatches: 0
 Query Match: 95.56% Indels: 0
 DB: 6 Gaps: 0

US-10-725-373-4 (1-9) x ABV87551 (1-409)

QY 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
 DB 241 TACCTTTCGGAGCGAACCCTCAACCTC 215

RESULT 15
 ABV89100
 ID ABV89100 standard; cDNA; 409 BP.

XX AC ABV89100;
 XX DT 13-DEC-2002 (first entry)
 XX DE Human colon cancer related cDNA SEQ ID NO 2411.
 XX KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 XX SS.
 XX OS Homo sapiens.
 XX PN WO200258534-A2.
 XX PD 01-AUG-2002.
 XX PF 16-NOV-2001; 2001WO-US043704.
 XX PR 20-NOV-2000; 2000US-0252222P.
 XX PR 06-FEB-2001; 2001US-0267011P.
 XX PR 28-MAR-2001; 2001US-0279670P.
 XX PR 10-JUL-2001; 2001US-0304037P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
 XX WPI; 2002-608400/65.

PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer.

PS Claim 1; SEQ ID NO 2411; 266pp + Sequence Listing; English.

CC The invention relates to a human colon tumour expressed polynucleotide
 CC (i) encoding a polypeptide (ii), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 409 BP; 103 A; 131 C; 76 G; 99 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 409
Score: 2.58 Length: 409
Percent Similarity: 43.00 Matches: 8
Best Local Similarity: 100.00% Conservative: 1
Query Match: 88.89% Mismatches: 0
DB: 95.56% Indels: 0
6 Gaps: 0

US-10-725-373-4 (1-9) x ABV89100 (1-409)

Qy 1 TyrLeuSerGlyAlaAsnIleAsnLeu 9
Db 169 TACCTTCGGGAGCGAACCTCAACCTC 195

Search completed: December 6, 2005, 16:29:54
Job time : 368.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 15:17:16 ; Search time 2995.75 Seconds
(without alignments)
170.772 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACLN 9

Scoring table: BLOSUM62
Ygapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5893141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODE=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10725373/runat_01122005_114445_21447/app_query.fasta_1.796
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373@cgn_1_1_5655@runat_01122005_114445_21447 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_ste.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	27	6	BD131679 Carcinoem
2	48	100.0	27	6	CS089183 Sequence
3	48	100.0	27	6	AR560608 Sequence

4	95.8	175953	9	AC154525	
5	95.8	252143	14	AC095437	
6	95.8	297470	14	AC125817	
c	7	91.7	163584	8	AC008056
8	91.7	250810	14	AC103486	
c	9	91.7	268281	14	AC124920
c	10	91.7	274993	14	AC133403
c	11	91.7	277735	14	CR753825
c	12	89.6	145377	14	AL645535
c	13	89.6	161363	9	AL645535
c	14	89.6	164991	8	AC092023
c	15	89.6	171347	8	AC099776
c	16	87.5	95663	8	AC010247
c	17	87.5	112428	15	AC122164
c	18	87.5	114109	8	AP002456
c	19	87.5	130442	6	Q0870468
c	20	87.5	161547	8	AP001929
c	21	87.5	167108	14	AC068283
c	22	87.5	189356	14	AC117866
c	23	87.5	203068	9	AL807755
c	24	87.5	207814	14	AC166061
c	25	87.5	213572	14	AC105567
c	26	87.5	246624	14	AC125751
c	27	87.5	349980	6	Q0869981
c	28	85.4	647	8	AF020692
c	29	85.4	1884	6	AX513524
c	30	85.4	1899	6	AR359364
c	31	85.4	1899	6	AX098631
c	32	85.4	1899	6	AX513520
c	33	85.4	1899	6	AX513522
c	34	85.4	1902	1	AF441855
c	35	85.4	1902	1	AY297091
c	36	85.4	1902	6	AR107058
c	37	85.4	1902	6	AR107059
c	38	85.4	1902	6	BD082244
c	39	85.4	1902	6	BD082245
c	40	85.4	1912	1	BACCR1B2
c	41	85.4	1912	6	AR260589
c	42	85.4	2281	1	AF164666
c	43	85.4	2610	2	AK173522
c	44	85.4	2765	5	AF403114
c	45	85.4	2812	5	AF403116

ALIGNMENTS

RESULT 1
BD131679
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BD131679
Carcinoembryonic antigen (CEA) agonist and antagonist peptides.
BD131679
BD131679.1 GI:23226624
JP 2002500002-A/5.
Homo sapiens (human)
Homo sapiens

27 bp
DNA
linear
PAT 18-SEP-2002

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

1 (bases 1 to 27)
Schlomo, J., Barzaga, E. and Zaremba, S.
Carcinoembryonic antigen (CEA) agonist and antagonist peptides
Patent: JP 2002500002-A 5 08-JAN-2002;
THE UNITED STATES OF AMERICA
OS Homo sapiens (human)
PN JP 2002500002-A/5
PD 08-JAN-2002
PF 22-SEP-1998 JP 2000516030
PR 10-OCT-1997 US 60/061589

PI JEFFREY SCHLOM, ELENE BARZAGA, SAM ZAREMBA
PC C12N15/09, A61K38/00, A61K45/00, A61P35/00, A61P37/02,
A61P43/00,
PC C07K14/705, C07K16/28, C12N5/10, C12N5/00, A61K37/02, C12N5/00 CC
Carcinoembryonic antigen (CEA) agonist and antagonist peptides FH

AC154525 Mus muscu
AC095437 Rattus no
AC125817 Rattus no
AC008056 Homo sapi
AC103486 Rattus no
AC124920 Rattus no
AC133403 Rattus no
CR753825 Homo sapi
AC158032 Bos tauru
AL645535 Mouse DNA
AL645535 Mouse DNA
AC092023 Homo sapi
AC099776 Homo sapi
AC010247 Homo sapi
AC122164 Medica90
AP002456 Homo sapi
Q0870468 Sequence
AP001929 Homo sapi
AC068283 Homo sapi
AC117866 Rattus no
AL807755 Mouse DNA
AC160661 Mus muscu
AC105567 Rattus no
AC125751 Rattus no
Q0869981 Sequence
AF020692 Saginus
AX513524 Sequence
AR359364 Sequence
AX098631 Sequence
AX513520 Sequence
AX513522 Sequence
AF441855 Bacillus
AY297091 Bacillus
AR107058 Sequence
AR107059 Sequence
BD082244 Bacillus
BD082245 Bacillus
BACCR1B2
M3724 B. thuringie
AR260589 Sequence
AF164666 Bacillus
AK173522 Ciona int
AF403114 Gallus ga
AF403116 Gallus ga

```

Key source Location/Qualifiers
FT source 1..27
FT Location/Qualifiers
FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.093 Length: 27
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-5 (1-9) x BD131679 (1-27)
QY 1 TyrLeuSerGlyAlaCysLeuAenLeu 9
DB 1 TACCTTCGGGAGCGTGCTCAACCTC 27

RESULT 2
LOCUS CS089183 27 bp DNA linear PAT 25-MAY-2005
DEFINITION Sequence 12 from Patent EP1447414.
ACCESSION CS089183
VERSION CS089183.1 GI:66714460
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Schlom, J., Salazar, M.E. and Zaremba, S.
TITLE Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
JOURNAL Patent: EP 1447414-A 12 18-AUG-2004;
Department of Health and Human Services (US)
FEATURES
source 1..27
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0.093 Length: 27
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-5 (1-9) x CS089183 (1-27)
QY 1 TyrLeuSerGlyAlaCysLeuAenLeu 9
DB 1 TACCTTCGGGAGCGTGCTCAACCTC 27

RESULT 3
LOCUS AR560608 27 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 12 from patent US 6756038.
ACCESSION AR560608
VERSION AR560608.1 GI:53972929
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 27)

Schlom, J., Barzaga, E. and Zaremba, S.
Agonist and antagonist peptides of carcinoembryonic antigen (CEA)
Patent: US 6756038-A 12 29-JUN-2004;
The United States of America as represented by the Department of
Health and Human Services; Washington, DC;
MOX;
FEATURES
source 1..27
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0.093 Length: 27
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-725-373-5 (1-9) x AR560608 (1-27)
QY 1 TyrLeuSerGlyAlaCysLeuAenLeu 9
DB 1 TACCTTCGGGAGCGTGCTCAACCTC 27

RESULT 4
LOCUS AC154525 175953 bp DNA linear ROD 30-JUL-2005
DEFINITION Mus musculus chromosome 14 clone RP24-6819, complete sequence.
ACCESSION AC154525 AC105170
VERSION AC154525.2 GI:71533348
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 175953)
AUTHORS Wilson, R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175953)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 175953)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 30, 2005 this sequence version replaced gi:56900151.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BB0068109
Drafting center: WIBR
-----
Location/Qualifiers
1..175953
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP24-6819"

FEATURES
source 1..175953
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP24-6819"

ORIGIN
Alignment Scores:
```

```

Pred. No.: 1.25e-03 Length: 175953
Score: 46.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 95.83% Indels: 0
DB: 9 Gaps: 0

US-10-725-373-5 (1-9) x AC154525 (1-175953)

OY 1 TyLeuserGlyAlaCysLeuAenLeu 9
Db 12406 TACCTAAGTGAGCATGATTAACCTT 12432

RESULT 5
AC095437 252143 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-4H9, *** SEQUENCING IN PROGRESS ***,
DEFINITION 2 unordered pieces.
AC095437
AC095437.6 GI:30467826
VERSION HTG: HTGS PHAS1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
1 (bases 1 to 252143)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guevara,W.,
Guevarra,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jacksone,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheua,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milobavjevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrellie,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

```

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 252143)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 252143)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24941147.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCHA
Center clone name: CH230-4H9
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 229361 bases at least Q40
Consensus quality: 231822 bases at least Q30
Consensus quality: 233149 bases at least Q20
Estimated insert size: 249796; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 246919: contig of 246919 bp in length
* 246920 252143: gap of unknown length
* 247020 252143: contig of 5124 bp in length.
FEATURES
source
1. 252143
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4H9"
/complement(2228..2885)
/note="clone boundary
clone_end:Sp6
site:EcoRI
end sequence:BH306797"
246920..247019
/estimated_length=unknown
misc_feature
247020..248193

```

/note="wgs_contig"

ORIGIN

Alignment Scores: 1.78e+03 Length: 252143
 Pred. No.: 46.00 Matches: 8
 Score: 100.00% Conservative: 1
 Percent Similarity: 88.89% Mismatches: 0
 Best Local Similarity: 88.89% Indels: 0
 Query Match: 95.83% Gaps: 0
 DB: 14

US-10-725-373-5 (1-9) x AC095437 (1-252143)

QY 1 TyLeuserGlyAlaCysLeuAenLeu 9

Db 132289 TACCTAAGCGGAGCATGCTTAACCTT 132315

RESULT 6

AC125817/c

LOCUS

AC125817 Rattus norvegicus clone CH230-2N21, WORKING DRAFT SEQUENCE, 6 linear HTG 13-NOV-2002

DEFINITION

unordered pieces.

AC125817 AC125817.3 GI:24635796

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Muridae; Murinae; Rattus;

1 (bases 1 to 297470)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshewa, L., Louibeged, H., Lozado, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindarcne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwakoelameh, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,

Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D. D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G. and Gibbs, R. A.

Unpublished

2 (bases 1 to 297470)

Worley, K. C.

Direct Submission

Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 297470)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 6, 2002 this sequence version replaced gi:22773028.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDCE

Center clone name: CH230-2N21

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218036 bases at least Q40

Consensus quality: 220310 bases at least Q30

Consensus quality: 221663 bases at least Q20

Estimated insert size: 223274; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 282417: contig of 282417 bp in length

* 282418 282517: gap of unknown length

* 282518 284692: contig of 2175 bp in length

* 284693 284792: gap of unknown length

* 284793 285902: contig of 1110 bp in length

* 285903 286002: gap of unknown length

* 286003 287368: contig of 1365 bp in length

* 287369 287469: gap of unknown length

* 287469 289494: contig of 2026 bp in length

* 289495 289594: gap of unknown length

* 289595 297470: contig of 7876 bp in length.

* Location/Qualifiers

1. .297470

FEATURES

source

gap	282418. .282517
gap	/estimated_length=unknown
gap	284693. .284792
gap	/estimated_length=unknown
gap	285903. .286002
gap	/estimated_length=unknown
gap	287369. .287468
gap	/estimated_length=unknown
gap	289495. .289594
gap	/estimated_length=unknown

Alignment Scores:		
Pred. No.:	2.08e+03	Length:
Score:	46.00	Matches:
Percent Similarity:	100.00%	Conservative:
Best Local Similarity:	88.69%	Mismatches:
Query Match:	95.83%	Indels:
DB:	14	Gaps:
		0
		0
		297470

US-10-725-373-5 (1-9) x AC125817 (1-297470)

1 TvrLeuSerGlyAlaCysLeuAsnLeu 9

Db 256539 TACCTAAGCGGAGCATGCATTAAACCTT 256513

RESULT 7

AC008056	135584 bp	DNA	linear	PRI 20-NOV-1999
LOCUS				
DEFINITION	Chromosome 14 clone RP11-242P2 containing gene for			
	neurexin III gene, partial cds, complete sequence.			

REFERENCE AUTHORS	SOURCE	ORGANISM	HLG.
		Homo sapiens (human)	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;	
		Hominidae; Homo.	
REFERENCE		1 (bases 1 to 163584)	
AUTHORS		Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,	
		Blowen, S., Dorr, M., Dickhoff, R., Fleetwood, P., Harrison, G.,	
		James, R., Kaur, A., Madan, A., Owen, M.P., Ratcliffe, A., Shafer, T.	
		and Hood, L.	

Chemistry: Big Dye terminators and primers
Assembly program: Phrap: version 0.990399

FEATURES

source

Location/Qualifiers

source	1..163584	/organism="Homo sapiens"	
		/mol_type="Genomic DNA"	
		/db_xref="taxon:9606"	
		/chromosome="14"	
		/map="14q31"	
		/clone="RP11-242p2"	
		/clone_lib="RPCI human BAC library 11"	
		/note="This clone overlaps RP11-45G3, Accession AC012099 and CTD-2294M16, AC007058"	
unsure	1450..1620	/note="low quality data in microsatellite region"	
unsure	29640..29680	/note="low quality data"	
unsure	62720..62830	/note="low quality data"	
CDS	join(<100016..100187,133868..>134175)		
		/note="Intron-exon boundaries defined in relation to KIAA0743"	
		/codon_start=1	
		/product="neurexin III"	
		/protein_id="AA09143.1"	
		/db_xref="GI:6456145"	
		/translation="EQKIGVFNIGTVDISIKERTPVNDKYHVFRFRNGNATLQVDNPVNEYPTGRQITFNTQAIAGGDKGRLPQGLSLGYDGLKVLNAAEN NPNKINGSVRLVGEVPSILGTTQTTSMPPEMSTTWETTWTATTTRKNRSTASIQ"	
unsure	110940..111010	/note="low quality data"	
unsure	119050..119080	/note="low quality data"	
misc_feature	146567..163584		
variation		/note="overlap with CTD-2294M16, Accession AC007058"	
	146826	/note="242p2: g; 2294M16: a"	
		/replace="a"	
variation	147921	/note="242p2: c; 2294M16: a"	
		/replace="t"	
variation	148707	/note="242p2: a; 2294M16: g"	
		/replace="g"	
variation	151285	/note="242p2: a; 2294M16: g"	
		/replace="g"	
variation	151480	/note="242p2: t; 2294M16: c"	
		/replace="c"	
variation	152361	/note="242p2: a; 2294M16: g"	
		/replace="g"	
variation	152736	/note="242p2: t; 2294M16: c"	
		/replace="c"	
variation	152804	/note="242p2: a; 2294M16: c"	
		/replace="c"	
variation	152837	/note="242p2: c; 2294M16: t"	
		/replace="t"	
variation	154549	/note="242p2: a; 2294M16: g"	
		/replace="g"	
variation	155489..155491		
		/note="242p2: atg; 2294M16: a"	
		/replace="a"	
variation	155537	/note="242p2: a; 2294M16: atatg"	
		/replace="atatg"	

```

variation 155586..155588
/notes="242P2: cat: 2294M16: c"
/replaces="c"
variation 155648..155650
/notes="242P2: ata: 2294M16: atgtgtgtg"
/replaces="atgtgtgtg"
variation 155882
/notes="242P2: a: 2294M16: c"
/replaces="c"

ORIGIN
Alignment Scores:
Pred. No.: 3.17e+03 Length: 163584
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 8 Gaps: 0

US-10-725-373-5 (1-9) x AC008056 (1-163584)

QY 1 TyrlouSerGlyAlaCysLeuAan 8
DB 144884 TACCTTTCGTGTCATGTTTGAAT 144907

RESULT 8
AC103486 AC103486 250810 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-19A6, WORKING DRAFT SEQUENCE, 2
DEFINITION
AC103486
AC103486.6 GI:30520423
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 250810)
Muzny,D,Marie,,Metzker,M,Lee,,Abramson,S,,Adams,C,,Alder,J.,
Allen,C,,Allen,H,,Alsbrooks,S,,Amin,A,,Anguiano,D.,
Anyalebechi,V,,Aoyagi,A,,Ayodeji,M,,Baca,E,,Baden,H.,
Baldwin,D,,Bandaranaika,D,,Barber,M,,Barnstead,M,,Benahmed,F.,
Biswalo,K,,Blair,J.,Blankenburg,K,,Blyth,P.,Brown,M.,
Bryant,N.,Buhay,C.,Burch,P.,Burrell,K.,Calderon,E.,
Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,
Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,
Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Fraser,C.M.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,
Gebregorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gunnarane,P.,Haaland,W.,Hamil,C.,Hamilton,C.,Hamilton,K.,
Harvey,Y.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Hulyk,S.,Hume,J.,Idlebird,D.,Jackson,A.,
Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolivet,A.,
Karpaty,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,
Kowis,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,
Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,
Lorenshewa,L.,Louiseed,H.,Lozado,R.J.,Lu,X.,Ma,J.,
Maheshwari,M.,Mahindartne,M.,Mahmoud,M.,Malloy,K.,Mangum,A.,
Mangum,B.,Mapua,P.,Martin,K.,Martin,R.,Martinez,E.,
Mawhiney,S.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,
Milosavljevic,A.,Miner,G.,Mirja,E.,Montemayor,J.,Moore,S.,
Morgan,M.,Morris,K.,Morris,S.,Munidas,M.,Murphy,M.,Naif,L.,
Nankervis,C.,Neal,D.,Newton,N.,Nguyen,N.,Norrie,S.,
Nwaokemele,O.,Okwuonu,G.,Olapunsaagoo,A.,Pal,S.,Parks,K.,
Pasternak,S.,Paul,H.,Perez,A.,Perez,L.,Pfankoch,C.,
Plopper,F.,Poindexter,A.,Popovic,D.,Primus,E.,Pu,L.-L.,

```

FEATURES

```

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wlaczkyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 250810)
Worley,K.C.
Direct Submission
Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 250810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:24942663.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be whole
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHTA
Center clone name: CH230-19A6
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 243081 bases at least Q40
Consensus quality: 245135 bases at least Q30
Consensus quality: 246521 bases at least Q20
Estimated insert size: 254009; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one 'clone'
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 249098: contig of 249098 bp in length
* 249099 249198: gap of unknown length
* 249199 250810: contig of 1612 bp in length.
Location/Qualifiers

```

```

source
1. .250810
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-19A6"
misc_feature
1. .1599
/note="wgs contig"
misc_feature
246136. .249098
/note="wgs contig"
249098. .249198
gap
/estimated_length=unknown
ORIGIN
Alignment Scores:
Pred. No.: 4.79e+03 Length: 250810
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0
US-10-725-373-5 (1-9) x AC103486 (1-250810)
QY 1 TyrluSerGlyAlaCysIeuAen 8
|||||
Db 115119 TATCTCTCAGCGCGCTGTTGAAT 115142
|||||
RESULT 9
AC124920/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-228J18, *** SEQUENCING IN PROGRESS
** 2 unordered pieces.
AC124920
AC124920.3 GI:231196075
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
1 (bases 1 to 268281)
Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsebrook,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,C., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebrgeorgie,E., Geer,K., Gill,R., Grady,M., Guerra,T., Guevara,M.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,J.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulv,K., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Primus,E., Pu.L.L.,
Plopper,F., Poindexter,A., Popovic,D., Reeves,K., Regier,M.A., Reigh,R.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Riggs,F.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,U., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 268281)
Worley,K.C.
Direct Submission
Submitted (20-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 268281)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21703468.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYNG
Center clone name: CH230-228J18
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 235034 bases at least Q40
Consensus quality: 238312 bases at least Q30
Consensus quality: 240333 bases at least Q20
Estimated insert size: 255321; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 265363: contig of 265363 bp in length
* 265364 265463: gap of unknown length
* 265464 268281: contig of 2818 bp in length.
Location/Qualifiers
1. .268281
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
FEATURES
source

```

```

/db_xref="taxon:10116"
/clone="CH230-228J18"
1. .1201
/note="wgs_contig"
263541. .265363
/note="wgs_contig"
265364. .265463
/estimated_length=unknown
265464. .266507
/note="wgs_contig"

ORIGIN
Alignment Scores:
Pred. No.:      5.12e+03      Length:      268281
Score:          44.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      91.67%      Indels:      0
DB:              14      Gaps:      0

US-10-725-373-5 (1-9) x AC124920 (1-268281)

QY 1 TyrlouSerGlyAlaCysLeuAen 8
DB 48802 TATCTCTCAGCGCCTGCTGAAT 48779

RESULT 10
AC133403/C
LOCUS AC133403 274993 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-31209, WORKING DRAFT SEQUENCE.
ACCESSION AC133403
VERSION AC133403.2 GI:25138990
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Rattus.
1 (bases 1 to 274993)
Muzny,D,Marie, Metzker,M,Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allien,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Kocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Greggeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Guarinate,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,I., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Louisegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mavhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelmele,O., Okwionu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,

```

FEATURES
source

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 274993)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 274993)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:2795083.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCBN
Center clone name: CH230-31209
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 263872 bases at least Q40
Consensus quality: 266813 bases at least Q30
Consensus quality: 268369 bases at least Q20
Estimated insert size: 272879; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one 'clone'.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 274993: contig of 274993 bp in length.
Location/Qualifiers
1. .274993

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-31209"
1. .1185
/note="wgs end_extension
clone_end:T7"
/notes="clone_boundary
clone_end:T7"
site:
end sequence:BZ156831"
complement(267606. .268445)
/notes="clone_boundary
clone_end:Sp6"
site:
end sequence:BZ156848"
273381. .274993
/notes="wgs end_extension
clone_end:Sp6"

ORIGIN

Alignment Scores:
Pred. No.: 5.24e+03 Length: 274993
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-5 (1-9) x AC133403 (1-274993)

QY 1 TTTLeuSerGlyAlaCysLeuAan 8

Db 152467 TATCTCTCAGGCGCCTGCTTGAAT 152444

RESULT 11

CR753825 277735 bp DNA linear HTG 19-MAR-2005
LOCUS
DEFINITION Homo sapiens chromosome 6 clone DASS-113F17, 32 unordered pieces.
ACCESSION
CR753825
VERSION
CR753825.3 GI:54260993
KEYWORDS
HTG; HTGS PHASE1; HTGS_CANCELLED.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE

1 (bases 1 to 277735)
Sims, S.
Direct Submission
Submitted (11-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 15, 2004 this sequence version replaced gi:51591719.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BSS113F17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 117368 bases at least Q40
Consensus quality: 117657 bases at least Q30
Consensus quality: 117840 bases at least Q20
Insert size: 274635; sum-of-contigs
Insert size: 127555; 13.8% error; agarose-fp
Quality coverage: 3.03x in Q20 bases; sum-of-contigs Quality
coverage: 6.53x in Q20 bases; agarose-fp

FEATURES
source

* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
3869: contig of 3869 bp in length
3870: gap of 100 bp
3970: contig of 22235 bp in length
26204: gap of 100 bp
26205: contig of 4000 bp in length
26305: gap of 100 bp
30304: contig of 4000 bp in length
30404: gap of 100 bp
34404: contig of 4000 bp in length
34504: gap of 100 bp
34505: contig of 4000 bp in length
38504: gap of 100 bp
38505: contig of 4000 bp in length
42604: gap of 100 bp
42605: contig of 4000 bp in length
42704: gap of 100 bp
46704: contig of 4000 bp in length
46804: gap of 100 bp
50804: contig of 4000 bp in length
50904: gap of 100 bp
50905: contig of 4000 bp in length
54904: gap of 100 bp
54905: contig of 4000 bp in length
55004: gap of 100 bp
55005: contig of 4000 bp in length
59104: gap of 100 bp
59105: contig of 4000 bp in length
63104: gap of 100 bp
63105: contig of 4000 bp in length
67204: gap of 100 bp
67205: contig of 4000 bp in length
71304: gap of 100 bp
71305: contig of 4000 bp in length
75404: gap of 100 bp
75405: contig of 4000 bp in length
79504: gap of 100 bp
79505: contig of 4000 bp in length
83604: gap of 100 bp
83605: contig of 4000 bp in length
87704: gap of 100 bp
87705: contig of 4000 bp in length
87804: gap of 100 bp
87805: contig of 4000 bp in length
91804: gap of 100 bp
91805: contig of 4000 bp in length
95904: gap of 100 bp
95905: contig of 4000 bp in length
96004: gap of 100 bp
96005: contig of 4000 bp in length
100004: gap of 100 bp
100005: contig of 4000 bp in length
104104: gap of 100 bp
104105: contig of 4000 bp in length
104204: gap of 100 bp
104205: contig of 4000 bp in length
108204: gap of 100 bp
108205: contig of 4000 bp in length
112304: gap of 100 bp
112305: contig of 4000 bp in length
116404: gap of 100 bp
116405: contig of 4000 bp in length
120504: gap of 100 bp
120505: contig of 4000 bp in length
124604: gap of 100 bp
124605: contig of 4000 bp in length
124704: gap of 100 bp
124705: contig of 4000 bp in length
128704: gap of 100 bp
128705: contig of 4000 bp in length
128804: gap of 100 bp
128805: contig of 4000 bp in length
132804: gap of 100 bp
132805: contig of 4000 bp in length
136904: gap of 100 bp
136905: contig of 4000 bp in length
141004: gap of 100 bp
141005: contig of 4000 bp in length
145104: gap of 100 bp
145105: contig of 4000 bp in length
145204: gap of 100 bp
145205: contig of 132531 bp in length.

Location/Qualifiers
1. .277735

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="DNA-arts-BAC.1-SSTO.1"
/clone="DASS-113F17"
1. .3869
/note="assembly_fragment:00042"
3970. .26204
/note="assembly_fragment:01266"
28305. .30304
/note="assembly_fragment:01274"
30405. .34404
/note="assembly_fragment:01275"
34505. .38504
/note="assembly_fragment:01276"
38605. .42604
/note="assembly_fragment:01277"
42705. .46704
/note="assembly_fragment:01278"
46805. .50804
/note="assembly_fragment:01279"
50905. .54904
/note="assembly_fragment:01280"
55005. .59004
/note="assembly_fragment:01281"
59105. .63104
/note="assembly_fragment:01282"
63205. .67204
/note="assembly_fragment:01283"
67305. .71304
/note="assembly_fragment:01284"
71405. .75404
/note="assembly_fragment:01285"
75505. .79504
/note="assembly_fragment:01286"
79605. .83604
/note="assembly_fragment:01287"
83705. .87704
/note="assembly_fragment:01288"
87805. .91804
/note="assembly_fragment:01289"
91905. .95904
/note="assembly_fragment:01290"
96005. .100004
/note="assembly_fragment:01291"
100105. .104104
/note="assembly_fragment:01292"
104205. .108204
/note="assembly_fragment:01293"
108305. .112304
/note="assembly_fragment:01294"
112405. .116404
/note="assembly_fragment:01295"
116505. .120504
/note="assembly_fragment:01296"
120605. .124604
/note="assembly_fragment:01297"
124705. .128704
/note="assembly_fragment:01298"
128805. .132804
/note="assembly_fragment:01299"
132905. .136904
/note="assembly_fragment:01300"
137005. .141004
/note="assembly_fragment:01301"
141105. .145104
/note="assembly_fragment:01302"
145205. .277735
/note="assembly_fragment:01303"
```

ORIGIN

Alignment Scores:

```
Pred. No.: 5.29e+03 Length: 277735
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-5 (1-9) x CR753825 (1-277735)

QY 1 TyrLeuSerClyAlaCysLeuAen 8
Db 174029 TACCTTTCGTGGTCGATGTTGAAT 174052

RESULT 12
AC158032/c
LOCUS AC158032 145377 bp DNA linear HTG 01-JUL-2005
DEFINITION Bos taurus clone CH240-145M23, WORKING DRAFT SEQUENCE, 3 unordered
pieces.
AC158032
VERSION AC158032.2 GI:68301101
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 145377)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davilla, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kwis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J.,
Lorensuhera, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
```

Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 145377)
 Worley, K.C.
 Direct Submission
 Submitted (04-MAR-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 145377)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jun 29, 2005 this sequence version replaced gi:60498804.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: FIH
 Center clone name: CH240-145M23
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 144398 bases at least Q40
 Consensus quality: 144751 bases at least Q30
 Consensus quality: 145038 bases at least Q20
 Estimated insert size: 146868; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 81420: contig of 81420 bp in length
 * 81421 81470: gap of 50 bp
 * 81471 144235: contig of 62765 bp in length
 * 144236 144335: gap of unknown length
 * 144336 145377: contig of 1042 bp in length.

FEATURES

Location/Qualifiers
 1. .145377
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="CH240-145M23"
 81421..81470
 /estimated_length=50
 144236..144335
 /estimated_length=unknown

ORIGIN

Alignment Scores:
 Pred. No.: 4.65e+03 Length: 145377
 Score: 43.00 Matches: 8
 Percent Similarity: 88.89% Conservative: 0
 Best Local Similarity: 88.89% Mismatches: 1
 Query Match: 89.58% Indels: 0
 DB: 14 Gaps: 0

US-10-725-373-5 (1-9) x AC158032 (1-145377)

QY 1 TyrlusSerGlyAlaCysLeuAsnLeu 9
 Db 15475 TATTGTCTGGCTGTGCTAAATCTT 15449

RESULT 13

LOCUS AL645535/c 161363 bp DNA linear ROD 09-AUG-2002
 DEFINITION Mouse DNA sequence from clone RP23-145G23 on chromosome 11,
 complete sequence.

ACCESSION AL645535

VERSION AL645535.16 GI:22204296

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 161363)

AUTHORS Bates, K.

TITLE Direct Submission

JOURNAL

Submitted (31-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 11, 2002 this sequence version replaced gi:22003119.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/c_elegans/wormpep RP23-145G23 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

Location/Qualifiers
 1. .161363
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-145G23"
 /clone_lib="RPCI-23"

ORIGIN

Alignment Scores:

Pred. No.: 5.15e+03 Length: 161363
Score: 43.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 89.58% Indels: 0
DB: 9 Gaps: 0

US-10-725-373-5 (1-9) x AL645535 (1-161363)

Qy 1 TyrlouSerGlyAlaCysleuhenleu 9
Db 142530 TATTTATCTGGACTGCTTGAACCTT 142504

RESULT 14

AC092023 AC092023 164991 bp DNA linear PRI 29-MAR-2002
LOCUS Homo sapiens chromosome 3 clone RP11-13C14, complete sequence.

DEFINITION AC092023 AC068300
ACCESSION AC092023.2 GI:19807852

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1 (bases 1 to 164991)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

TITLE Direct Submission

REFERENCE 2 (bases 1 to 164991)

AUTHORS Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and
Haugen, E.D.

TITLE Direct Submission

JOURNAL Submitted (15-JUN-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

REFERENCE 3 (bases 1 to 164991)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.

TITLE Direct Submission

JOURNAL Submitted (29-MAR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

COMMENT On Mar 29, 2002 this sequence version replaced gi:14456651.

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Contact: uwgctgs@u.washington.edu

Drafting Center: BCM

----- Project Information

Center project name: chr-3

Center clone name: RP11-13C14 (bc0121)

----- Summary Statistics

Sequencing vector: M13; L08821; 51% of reads

Chemistry: Dye-terminator ET; 45% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 164525 bases at least Q40

Consensus quality: 164952 bases at least Q30

Consensus quality: 164982 bases at least Q20

Insert size: 164991; sum-of-contigs

Quality coverage: 7.3x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': Mapping in progress

3': RP11-28013 (UWGC:bc0336) AC0999776, 46806-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

HindIII EcoRI BglII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----
1583	1636	8696	8776	4390	4260
-----	-----	-----	-----	-----	-----
6382	6557	6	<800	2067	2068
-----	-----	-----	-----	-----	-----
512	<800	25	<800	5379	5535
-----	-----	-----	-----	-----	-----
449	<800	425	<800	1330	1285
-----	-----	-----	-----	-----	-----
5430	5264	901	902	5365	5316
-----	-----	-----	-----	-----	-----
3121	3135	688	<800	10122	10079
-----	-----	-----	-----	-----	-----
3145	3135	1163	1141	6473	6360
-----	-----	-----	-----	-----	-----
2066	2028	3545	3462	6074	6073
-----	-----	-----	-----	-----	-----
3615	3651	8104	8046	744	<800
-----	-----	-----	-----	-----	-----
179	<800	4625	4601	2983	3022
-----	-----	-----	-----	-----	-----
1171	1109	901	902	3025	3022
-----	-----	-----	-----	-----	-----
678	<800	6384	6326	6310	6360
-----	-----	-----	-----	-----	-----
633	<800	1265	1264	1918	1900
-----	-----	-----	-----	-----	-----
342	<800	1493	1472	109	<800
-----	-----	-----	-----	-----	-----
2338	2307	222	<800	3209	3220
-----	-----	-----	-----	-----	-----
1836	1765	2105	2132	497	<800
-----	-----	-----	-----	-----	-----
204	<800	2983	2914	32	<800
-----	-----	-----	-----	-----	-----
755	747	3094	3142	2110	2068
-----	-----	-----	-----	-----	-----
5308	5264	6997	6938	1094	1082
-----	-----	-----	-----	-----	-----
1639	1636	1765	1756	70	<800
-----	-----	-----	-----	-----	-----
3582	3651	4493	4464	2872	2879
-----	-----	-----	-----	-----	-----
103	<800	2165	2132	1054	1082

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular HAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII				BglII				EcoRI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1748	1716	5104	4922	8696	8914						
-----	-----	-----	-----	-----	-----						
6382	6590	2067	2079	6	<800						
-----	-----	-----	-----	-----	-----						
512	<800	4720	4922	1720	1746						
-----	-----	-----	-----	-----	-----						
449	<800	3762	3779	5311	5156						
-----	-----	-----	-----	-----	-----						
5459	5330	9341	9234	278	<800						
-----	-----	-----	-----	-----	-----						
471	<800	3566	3583	1789	1746						
-----	-----	-----	-----	-----	-----						
4419	4423	5988	6069	20	<800						
-----	-----	-----	-----	-----	-----						
376	<800	1888	1875	5487	5395						
-----	-----	-----	-----	-----	-----						
2049	2022	13116	12878	2109	2110						
-----	-----	-----	-----	-----	-----						
7643	7617	2246	2250	3130	3320						
-----	-----	-----	-----	-----	-----						
1987	2022	2091	2079	2844	2875						
-----	-----	-----	-----	-----	-----						
5893	6014	97	<800	4968	5156						
-----	-----	-----	-----	-----	-----						
6633	6590	2204	2250	2663	2704						
-----	-----	-----	-----	-----	-----						
6647	6590	8350	8230	6354	6301						
-----	-----	-----	-----	-----	-----						
1828	1884	7183	7228	3923	3873						
-----	-----	-----	-----	-----	-----						
3141	3131	2563	2631	1583	1637						
-----	-----	-----	-----	-----	-----						
4271	4244	3291	3346	1388	1375						
-----	-----	-----	-----	-----	-----						
7989	7826	4260	4231	1517	1490						
-----	-----	-----	-----	-----	-----						
1083	1077	1506	1447	1683	1637						
-----	-----	-----	-----	-----	-----						
3123	3131	2415	2453	2529	2567						
-----	-----	-----	-----	-----	-----						
1362	1328	2966	2985	567	<800						
-----	-----	-----	-----	-----	-----						
5413	5330	6959	6931	1127	1111						
-----	-----	-----	-----	-----	-----						

Alignment Scores:

Pred. No.: 5.46e+03
Score: 43.00
Percent Similarity: 100.00%
Best Local Similarity: 77.78%

Length: 171347
Matches: 7
Conservative: 2
Mismatches: 0

1890	1884	747	757	452	<800
-----	-----	-----	-----	-----	-----
3627	3607	1624	1594	1516	1490
-----	-----	-----	-----	-----	-----
55	<800	4850	4922	4657	4576
-----	-----	-----	-----	-----	-----
2336	2319	1663	1594	125	<800
-----	-----	-----	-----	-----	-----
6686	6590	7719	7669	11257	11227
-----	-----	-----	-----	-----	-----
1886	1884	281	<800	5162	5156
-----	-----	-----	-----	-----	-----
2442	2454	3112	3171	574	<800
-----	-----	-----	-----	-----	-----
254	<800	8712	8596	562	<800
-----	-----	-----	-----	-----	-----
7392	7364	488	<800	8296	8219
-----	-----	-----	-----	-----	-----
224	<800	760	757	3316	3423
-----	-----	-----	-----	-----	-----
476	<800	2042	2079	746	761
-----	-----	-----	-----	-----	-----
5240	5150	2396	2453	426	<800
-----	-----	-----	-----	-----	-----
334	<800	20802	21047	9062	8914
-----	-----	-----	-----	-----	-----
4224	4244	3471	3475	4685	4576
-----	-----	-----	-----	-----	-----
2487	2454	1318	1276	3832	3873
-----	-----	-----	-----	-----	-----
2604	2654	2049	2079	3973	3873
-----	-----	-----	-----	-----	-----
5186	5150	3761	3779	5273	5156
-----	-----	-----	-----	-----	-----
1256	1227	4354	4400	1538	1490
-----	-----	-----	-----	-----	-----
161	<800	9087	8991	2211	2201
-----	-----	-----	-----	-----	-----
4391	4423	1330	1276	3495	3423
-----	-----	-----	-----	-----	-----
1892	1884	2193	2250	1680	1637
-----	-----	-----	-----	-----	-----
541	<800	1604	1594	8426	8219
-----	-----	-----	-----	-----	-----
3674	3607			608	<800
-----	-----	-----	-----	-----	-----
2525	2562			140	<800
-----	-----	-----	-----	-----	-----
5552	5489			3404	3423
-----	-----	-----	-----	-----	-----
6957	6899			3010	3025
-----	-----	-----	-----	-----	-----
873	886			1403	1375
-----	-----	-----	-----	-----	-----
6052	6014			5673	5552
-----	-----	-----	-----	-----	-----
347	<800			2391	2436
-----	-----	-----	-----	-----	-----
6135	6014			4956	4919
-----	-----	-----	-----	-----	-----
511	<800			12932	13091
-----	-----	-----	-----	-----	-----
3873	3821			1522	1490
-----	-----	-----	-----	-----	-----
487	<800			3061	3174
-----	-----	-----	-----	-----	-----
4085	4063				

Query Match: 89.58% Indels: 0
DB: 8 Gaps: 0

US-10-725-373-5 (1-9) x AC099776 (1-171347)

Qy 1 TyrLeuSerGlyAlaCysLeuLeu 9
Db 166976 TATCTGAGTGGTTCATGTATGAACCTC 166950

Search completed: December 6, 2005, 19:55:30
Job time : 3229.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2005, 10:17:58 ; Search time 367.25 Seconds
(without alignments)
163.328 Million cell updates/sec

Title: US-10-725-373-5
Perfect score: 48
Sequence: 1 YLSGACNLN 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPO2.spool/US10725373/runat_01122005_114444_21420/app_query.fasta_1.796
-DB=N Geneseg -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNIT5=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10725373 @CGN_1_1244 @runat_01122005_114444_21420 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseg_21.*

- 1: Geneseg1980s.*
- 2: Geneseg1990s.*
- 3: Geneseg2000s.*
- 4: Geneseg2001as.*
- 5: Geneseg2001bs.*
- 6: Geneseg2002as.*
- 7: Geneseg2002bs.*
- 8: Geneseg2003as.*
- 9: Geneseg2003bs.*
- 10: Geneseg2003cs.*
- 11: Geneseg2003ds.*
- 12: Geneseg2004as.*
- 13: Geneseg2004bs.*
- 14: Geneseg2005as.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	87.5	110000	13 ABD32780_3	Continuation (4 of
2	41	85.4	1176	13 ADX63202_3	Adx63202 Plant ful
3	41	85.4	1884	10 ADF60975	Adf60975 B. thurin
4	41	85.4	1899	4 AAS02464	Aas02464 B. thurin

5	41	85.4	1899	10 ADF60973	Adf60973 B. thurin
6	41	85.4	1899	10 ADF60971	Adf60971 B. thurin
7	41	85.4	1902	2 AAV52612	AAV52612 Nucleotid
8	41	85.4	1902	2 AAV52611	AAV52611 Nucleotid
9	41	85.4	1912	3 AAA15566	AAA15566 Cry2Ab de
10	41	85.4	2924	2 AAQ71027	AAQ71027 CryIIB ge
11	39	81.2	822	4 AAH52607	AAH52607 S. epider
12	39	81.2	1032	13 AD545359	AD545359 Bacterial
13	39	81.2	1384	3 AAC74312	AAC74312 Human sec
14	39	81.2	1717	14 AEB67311	Aeb67311 Rice geno
15	39	81.2	2950	4 AAH53985	AAH53985 S. epider
16	39	81.2	3760	4 AAH54665	AAH54665 S. epider
17	39	81.2	41202	12 ADQ97382	ADQ97382 Mouse can
18	39	81.2	80332	11 ACN44842	ACN44842 Human gen
19	38	79.2	374	8 ABZ19661	ABZ19661 Group III
20	38	79.2	404	13 ADX35484	Adx35484 Plant ful
21	38	79.2	476	13 ACF85844	ACF85844 Human SIR
22	38	79.2	610	13 ADQ57642	Adq57642 Novel can
23	38	79.2	848	12 ADJ67429	Adj67429 Human ova
24	38	79.2	862	12 ADJ67428	Adj67428 Human ova
25	38	79.2	1040	13 ADT17488	Adt17488 Plant CDN
26	38	79.2	1509	11 ABD00239	ABD00239 Klebsiell
27	38	79.2	2540	4 ABL26320	ABL26320 Drosophil
28	38	79.2	5002	12 ADQ64368	Adq64368 Novel hum
29	38	79.2	8801	4 ABL28642	ABL28642 Drosophil
30	38	79.2	10793	4 ABL26318	ABL26318 Drosophil
31	38	79.2	42999	6 ABK90832	ABK90832 Genomic D
32	38	79.2	51558	13 ACN37207	ACN37207 Human per
33	38	79.2	110000	6 ABX08336_04	Continuation (5 of
34	38	79.2	110000	12 ADJ25985_04	Continuation (5 of
35	38	79.2	110000	12 ADN97989_04	Continuation (5 of
36	38	79.2	110000	12 ADO50281_04	Continuation (5 of
37	38	79.2	110000	14 AEB85185_04	Continuation (5 of
38	38	79.2	114411	12 ADQ21090	Adq21090 Human sof
39	37	77.1	250	4 AAK69831	AAK69831 Human imm
40	37	77.1	300	4 AAK69830	AAK69830 Human imm
41	37	77.1	300	4 AAK57233	AAK57233 Human imm
42	37	77.1	354	6 ABN18754	ABn18754 Human ORF
43	37	77.1	375	8 ABZ18602	ABz18602 Group III
44	37	77.1	401	4 AAK96284	AAK96284 Human neu
45	37	77.1	401	4 AAK97777	AAK97777 Human neu

ALIGNMENTS

RESULT 1
ABD32780_3
Continuation (4 of 5) of ABD32780 from base 300001 (Human cancer-associated genomic DNA)
WP Sequence split into 5 fragments LOCUS ABD32780 Accession ABD32780

WP	Fragment Name	Begin	End
WP	ABD32780_0	1	110000
WP	ABD32780_1	100001	210000
WP	ABD32780_2	200001	310000
WP	ABD32780_3	300001	410000
WP	ABD32780_4	400001	430442

Alignment Scores:

Pred. No.: 3.15e+04
Score: 42.00
Percent Similarity: 88.89%
Best Local Similarity: 88.89%
Query Match: 87.50%
DB: 13

US-10-725-373-5 (1-9) x ABD32780_3 (1-110000)

Qy 1 TyrluSerGlyAlaCysLeuAsnLeu 9
Db 47157 TATCTGTCTGGCATTGCTTAACCTC 47183
RESULT 2
ADX63202/c
ID ADX63202 standard; cDNA; 1176 BP.

XX AC ADXG3202;
 XX DT 21-APR-2005 (first entry)
 XX DE Plant full length insert polynucleotide seqid 34045.
 XX KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content; gene; ss.
 XX OS Unidentified.
 XX PN US2004034888-A1.
 XX PD 19-FEB-2004.
 XX PF 28-APR-2003; 2003US-00425114.
 XX PR 06-MAY-1999; 99US-00304517.
 XX PR 05-NOV-2001; 2001US-00985678.
 XX PA (LIUJ/) LIU J.
 XX PA (ZHOU/) ZHOU Y.
 XX PA (KOVA/) KOVALIC D K.
 XX PA (SCRE/) SCREEN S E.
 XX PA (TABAS/) TABASKA J E.
 XX PA (CAOY/) CAO Y.
 XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 XX WPI; 2004-180133/17.
 XX PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.
 XX PS Claim 1; SEQ ID NO 34045; 15pp; English.
 XX CC The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in
 CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This sequence represents a plant full length insert
 CC polynucleotide that can be used in the recombinant DNA construct of the
 CC invention.

US-10-725-373-5 (1-9) x ADXG3202 (1-1176)
 QY 2 LeuSerGlyAlaCysLeuAsnLeu 9
 DB 651 CTCAGTGGCGCTTGCCCTCAATCTG 628
 RESULT 3
 ADF60975
 ID ADF60975 standard; DNA; 1884 BP.
 XX AC ADF60975;
 XX DT 12-FEB-2004 (first entry)
 XX DE B. thuringiensis Cry2Ag DNA.
 XX KW Cry2Ag; gene; ds; insecticidal protein; insecticide; insect control;
 KW insect damage; plant resistance.
 XX OS Bacillus thuringiensis.
 XX FH Key Location/Qualifiers
 FT CDS 1..1884
 FT /*tag= a
 FT /product= "B. thuringiensis Cry2Ag"
 XX PN US2003167517-A1.
 XX PD 04-SEP-2003.
 XX PF 09-JAN-2002; 2002US-00040906.
 XX PR 09-JAN-2001; 2001US-00331355.
 XX PA (ARNA/) ARNAUT G.
 XX PA (BOET/) BOETS A.
 XX PA (VANN/) VANNESTE S.
 XX PA (VRJE/) VAN RIE J.
 XX PA (VHOU/) VAN HOUDT S.
 XX PI Arnaut G, Boets A, Vanneste S, Van Rie J, Van Houdt S;
 XX WPI; 2003-898134/82.
 XX P-PSDB; ADF60976.
 XX PT New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
 PT protecting plants from insect damage, for controlling insects, or for
 PT rendering a plant resistant to an insect.
 XX PS Claim 7; SEQ ID NO 5; 32pp; English.
 XX CC The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
 CC and Cry2Ag. The invention also relates to nucleic acid sequences encoding
 CC the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
 CC comprising one of the nucleic acid sequences under the control of a plant
 CC -expressible promoter, plant cells, plants or seeds transformed to
 CC comprise the chimeric gene, a microorganism transformed to comprise the
 CC nucleic acid sequence, a process for rendering a plant resistant to an
 CC insect comprising transforming plant cells with the chimeric gene and
 CC regenerating transformed plants from the cells that are resistant to
 CC insects, and a method for controlling insects comprising expressing the
 CC Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
 CC Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
 CC proteins are useful for protecting plants from insect damage, for
 CC controlling insects or for rendering a plant resistant to an insect. This
 CC sequence represents DNA encoding the Cry2Ag protein of the invention.

XX SQ Sequence 1884 BP; 638 A; 292 C; 348 G; 606 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 465 Length: 1884
 Score: 41.00 Matches: 7

Percent Similarity: 88.89% Conservative: 1
 Best Local Similarity: 77.78% Mismatches: 1
 Query Match: 85.42% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-5 (1-9) x ADF60975 (1-1884)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
 DB 767 TATCTATCTGCTGCTGTTTAAATATC 793

RESULT 4
 AAS02464
 ID AAS02464 standard; DNA; 1899 BP.

AC AAS02464;
 XX 29-AUG-2001 (first entry)

DE B. thuringiensis DNA encoding a toxic crystal protein, CryET31.

KW Delta endotoxin; Lepidopteran-active; crystal protein; insecticide;
 transgenic plant; corn; wheat; soybean; oat; cotton; rice; sorghum;
 sugarcane; tomato; tobacco; kapok; flax; potato; barley; turf grass;
 pasture grass; berry; fruit; legume; vegetable; ornamental plant; shrub;
 cactus; tree cell; gypsy moth; looper; tobacco budworm; spruce budworm;
 cotton leaf perforator; CryET31; ds.

OS Bacillus thuringiensis.

XX Key Location/Qualifiers
 FH CDS 1..1899
 FT /*tag= a
 FT /product= "CryET31"

XX WO200119859-A2.

XX 22-MAR-2001.

PF 13-SEP-2000; 2000WO-US025361.

PR 15-SEP-1999; 99US-0153995P.

XX (MONS) MONSANTO CO.

PI Baum JA, Chu C, Donovan WP, Gilmer AJ, Rupar MJ;

XX WPI; 2001-281518/29.

DR P-PSDB; AAU02021.

XX Lepidopteran-active Bacillus thuringiensis delta-endotoxin polypeptides
 and the polynucleotides that encode them, useful for increasing the
 insect resistance of plant.

PS Claim 17; Page 99-102; 173pp; English.

XX The sequence encodes a B. thuringiensis Lepidopteran-active delta-
 endotoxin, crystal protein CryET31. The Lepidopteran-active B.
 thuringiensis delta-endotoxin polypeptides may be used as compositions
 that are applied to plant crops to protect them from insect damage. The
 polynucleotides may be used in the production of transgenic plants that
 express the insecticidal polypeptides and consequently have improved
 insect resistance compared to non-transformed plants. Monocotyledonous or
 dicotyledonous plants may be protected in this way, for example corn,
 wheat, soybean, oat, cotton, rice, rye, sorghum, sugarcane, tomato,
 tobacco, kapok, flax, potato, barley, turf grass, pasture grass, berry,
 fruit, legume, vegetable, ornamental plant, shrub, cactus and/or tree
 cell. A wide range of insects (e.g. gypsy moth, looper, tobacco budworm,
 cotton leaf perforator and spruce budworm) may be affected by application
 of the insecticidal polypeptides (full details given in specification)

XX Sequence 1899 BP; 619 A; 303 C; 356 G; 621 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 469 Length: 1899
 Score: 41.00 Matches: 7
 Percent Similarity: 88.89% Conservative: 1
 Best Local Similarity: 77.78% Mismatches: 1
 Query Match: 85.42% Indels: 0
 DB: 4 Gaps: 0

US-10-725-373-5 (1-9) x AAS02464 (1-1899)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
 DB 767 TATCTATCTGCTGCTGTTTAAATATC 793

RESULT 5

ADP60973

ID ADP60973 standard; DNA; 1899 BP.

XX ADP60973;

DT 12-FEB-2004 (first entry)

XX B. thuringiensis Cry2Af DNA.

XX Cry2Af; gene; ds; insecticidal protein; insecticide; insect control;
 insect damage; plant resistance.

XX Bacillus thuringiensis.

XX Key Location/Qualifiers
 FH CDS 1..1899
 FT /*tag= a
 FT /product= "B. thuringiensis Cry2Af"

XX US2003167517-A1.

XX 04-SEP-2003.

XX 09-JAN-2002; 2002US-00040906.

XX 09-JAN-2001; 2001US-00331355.

XX (ARNA/) ARNAUT G.

PA (BOET/) BOETS A.

PA (VANN/) VANNESTE S.

PA (VRIE/) VAN RIE J.

PA (VHOU/) VAN HOUDT S.

XX Arnaut G, Boets A, Vanneste S, Van Rie J, Van Houdt S;

XX WPI; 2003-898134/82.

DR P-PSDB; ADF60974.

XX New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
 protecting plants from insect damage, for controlling insects, or for
 rendering a plant resistant to an insect.

XX Claim 4; SEQ ID NO 3; 32pp; English.

XX The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
 and Cry2Ag. The invention also relates to nucleic acid sequences encoding
 the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
 comprising one of the nucleic acid sequences under the control of a plant
 -expressible promoter, plant cells, plants or seeds transformed to
 comprise the chimeric gene, a microorganism transformed to comprise the
 nucleic acid sequence, a process for rendering a plant resistant to an
 insect comprising transforming plant cells with the chimeric gene and
 regenerating transformed plants from the cells that are resistant to
 insects, and a method for controlling insects comprising expressing the
 Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
 Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
 proteins are useful for protecting plants from insect damage, for
 controlling insects or for rendering a plant resistant to an insect. This

CC sequence represents DNA encoding the Cry2Af protein of the invention.
 SQ Sequence 1899 BP; 624 A; 300 C; 350 G; 625 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 469 Length: 1899
 Score: 41.00 Matches: 7
 Percent Similarity: 88.89% Conservative: 1
 Best Local Similarity: 77.78% Mismatches: 1
 Query Match: 85.42% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-5 (1-9) x ADF60973 (1-1899)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9

DB 767 TATCTATCTGTCGTTGTTTAAATATC 793

RESULT 6

ID ADF60971 standard; DNA; 1899 BP.

XX ADF60971;

DT 12-FEB-2004 (first entry)

DE B. thuringiensis Cry2Ae DNA.

KW Cry2Ae; gene; ds; insecticidal protein; insecticide; insect control;
 insect damage; plant resistance.

OS Bacillus thuringiensis.

PH Key Location/Qualifiers

FT CDS 1..1899

FT /*tag= a

FT /product= "B. thuringiensis Cry2Ae"

XX US2003167517-A1.

XX 04-SEP-2003.

XX 09-JAN-2002; 2002US-00040906.

XX 09-JAN-2001; 2001US-00331355.

XX (ARNA/) ARNAUT G.

XX (BOET/) BOETS A.

XX (VANN/) VANNESTE S.

XX (VRIE/) VAN RIE J.

XX (VHOU/) VAN HOUT J.

XX Arnaut G, Boets A, Vanneste S, Van Rie J, Van Hout J;
 WPI: 2003-898134/82.

XX P-PSDB; ADF60972.

XX New Cry2Ae, Cry2Af and Cry2Ag insecticidal proteins, useful for
 protecting plants from insect damage, for controlling insects, or for
 rendering a plant resistant to an insect.

XX Claim 1; SEQ ID NO 1; 32pp; English.

XX The invention relates to insecticidal proteins, designated Cry2Ae, Cry2Af
 and Cry2Ag. The invention also relates to nucleic acid sequences encoding
 the insecticidal Cry2Ae, Cry2Af and Cry2Ag proteins, a chimeric gene
 comprising one of the nucleic acid sequences under the control of a plant
 -expressible promoter, plant cell, plants or seeds transformed to
 comprise the chimeric gene, a microorganism transformed to comprise the
 nucleic acid sequence, a process for rendering a plant resistant to an
 insect comprising transforming plant cells with the chimeric gene and
 regenerating transformed plants from the cells that are resistant to
 insects, and a method for controlling insects comprising expressing the

CC Cry2Ae, Cry2Af and Cry2Ag proteins in transformed plant cells. The
 CC Cry2Ae, Cry2Af and Cry2Ag proteins and nucleic acids encoding the
 CC proteins are useful for protecting plants from insect damage, for
 CC controlling insects or for rendering a plant resistant to an insect. This
 CC sequence represents DNA encoding the Cry2Ae protein of the invention.

SQ Sequence 1899 BP; 619 A; 303 C; 356 G; 621 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 469 Length: 1899
 Score: 41.00 Matches: 7
 Percent Similarity: 88.89% Conservative: 1
 Best Local Similarity: 77.78% Mismatches: 1
 Query Match: 85.42% Indels: 0
 DB: 10 Gaps: 0

US-10-725-373-5 (1-9) x ADF60971 (1-1899)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9

DB 767 TATCTATCTGTCGTTGTTTAAATATC 793

RESULT 7

AAV52612

ID AAV52612 standard; DNA; 1902 BP.

XX AAV52612;

AC AAV52612;

DT 02-DEC-1998 (first entry)

DE Nucleotide sequence of lepidoteran-active 8612 toxin.

KW 8612 toxin; PCR: primer; amplification; Bacillus thuringiensis; probe;
 lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
 Helicoverpa zea; hybridisation; ss.

XX Bacillus thuringiensis.

PH Key Location/Qualifiers

FT CDS 1..1902

FT /*tag= a

FT /product= "8612 toxin"

XX WO9840490-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US005081.

XX 13-MAR-1997; 97US-0040512P.

XX (MYCO) MYCOGEN CORP.

XX Schnepf HE, Narva KE, Muller-Cohn J;

XX WPI: 1998-506734/43.

XX P-PSDB; AAV57575.

XX New insecticidal Bacillus thuringiensis toxins - useful for controlling
 lepidopteran pests, especially Ostrinia nubilalis, Heliothis virescens
 and Helicoverpa zea.

XX Claim 12; Page 35-36; 50pp; English.

XX This is the nucleotide sequence of a novel Bacillus thuringiensis toxin
 used in the method of the invention, to control lepidopteran pests. The
 CC new toxins are useful as pesticides, especially for the control of
 CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
 CC polynucleotide coding sequences are useful for recombinant expression of
 CC the toxins and the primers, together with probes derived from the new
 CC sequences, are useful for the identification and characterisation of
 CC novel genes that encode pesticidal toxins

SQ Sequence 1902 BP; 637 A; 302 C; 338 G; 625 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 470 Length: 1902
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 2 Gaps: 0

US-10-725-373-5 (1-9) x AAV52612 (1-1902)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
DB 767 TATCTATCTGCTGCTGTTTAAATATC 793

RESULT 8

AAV52611
ID AAV52611 standard; DNA; 1902 BP.

XX AAV52611;

DT 02-DEC-1998 (first entry)

DE Nucleotide sequence of lepidoteran-active HD525 toxin.

KW HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
KW lepidoteran; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW Helicoverpa zea; hybridisation; ss.

OS Bacillus thuringiensis.

PH Key Location/Qualifiers
FT CDS 1..1902
FT /*tag= a
FT /product= "HD525 toxin"

PN WO9840490-A1.

PD 17-SEP-1998.

PF 13-MAR-1998; 98WO-US005081.

PR 13-MAR-1997; 97US-0040512P.

PA (MYCO) MYCOGEN CORP.

PI Schnepf HE, Narva KE, Muller-Cohn J;

DR WPI; 1998-506734/43.

DR P-PSDB; AAV75774.

XX New insecticidal Bacillus thuringiensis toxins - useful for controlling
PT lepidoteran pests, especially Ostrinia nubilalis, Heliothis virescens
PT and Helicoverpa zea.

PS Claim 17; Page 31-32; 50pp; English.

XX This is the nucleotide sequence of a novel Bacillus thuringiensis toxin
CC used in the method of the invention, to control lepidoteran pests. The
CC new toxins are useful as pesticides, especially for the control of
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC polynucleotide coding sequences are useful for recombinant expression of
CC the toxins and the primers, together with probes derived from the new
CC sequences, are useful for the identification and characterisation of
CC novel genes that encode pesticidal toxins

SQ Sequence 1902 BP; 633 A; 304 C; 337 G; 628 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 470 Length: 1902
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1

Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 2 Gaps: 0

US-10-725-373-5 (1-9) x AAV52611 (1-1902)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
DB 767 TATCTATCTGCTGCTGTTTAAATATC 793

RESULT 9

AAV5566
ID AAV5566 standard; DNA; 1912 BP.

XX AAV5566;

DT 28-JUL-2000 (first entry)

DE Cry2Ab delta-endotoxin gene.

XX Transgenic plant; insect resistance; cry2Ab delta-endotoxin; Coleopteran;
KW Lepidoteran; Dipteran; plasmid transit peptide; PTP; insecticidal;
KW plasmid targeting peptide; ds.

OS Bacillus thuringiensis.

PH Key Location/Qualifiers
FT CDS 1..1902
FT /*tag= a
FT /product= "cry2Ab delta-endotoxin"

PN WO200026371-A1.

PD 11-MAY-2000.

PF 04-NOV-1999; 99WO-US026086.

PR 04-NOV-1998; 98US-00186002.

PA (MONS) MONSANTO CO.

PI Corbin DR, Romano CP;

DR WPI; 2000-376130/32.

DR P-PSDB; AAY94260.

XX New method of expressing insecticidal proteins in plants transformed with
PT a Bacillus thuringiensis delta-endotoxin encoding gene resulting in
PT effective control of susceptible target pests.

PS Claim 12; Page 99; 104pp; English.

XX The present sequence is the cry2Ab delta-endotoxin gene. Delta-endotoxins
CC are produced by Bacillus thuringiensis during sporulation. These proteins
CC are toxic to certain species of insect e.g. Lepidoteran and Coleopteran
CC larvae. An insect-resistant transgenic plant has been constructed which
CC contains the present sequence. The cry2Ab gene would be transferred into
CC plants via expression vectors, which subsequently allow high expression
CC of the cry2Ab gene. The present sequence lacks Dipteran inhibitory
CC activity. Protection may be attained against insects such as Ostrinia
CC spp., Diatraea spp., Helicoverpa spp., and Spodoptera spp., in Zea mays;
CC Heliothis virescens, Helicoverpa spp., Pectinophora spp., in Gossypium
CC hirsutum; Anticarsia spp., Pseudoplusia spp., Epinotia spp., in Glycine
CC max; and Scirpophaga incertulas in Oryza sativa. Expression of the
CC present sequence by a plant cell produces a fusion protein comprising an
CC amino-terminal plasmid transit peptide (PTP) covalently linked to the
CC delta-endotoxin. The fusion protein functions to localise the delta-
CC endotoxin to a subcellular organelle or compartment

SQ Sequence 1912 BP; 627 A; 304 C; 351 G; 630 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 472 Length: 1912


```
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 3 Gaps: 0

US-10-725-373-5 (1-9) x AAA15566 (1-1912)
QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
Db 767 TAATCTATCGTGGTCGTTGTTAAATATC 793

RESULT 10
AAQ71027
ID AAQ71027 standard; DNA; 2924 BP.
XX
AC AAQ71027;
XX
DT 25-MAR-2003 (revised)
DT 27-MAR-1995 (first entry)
XX
DE CryIIB gene which encodes insecticidal crystal protein.
XX
KW CryIIA; CryIIB; CryC; P-2; CryBI; insecticidal protein crystal;
KW lepidoptera; environmental insecticide; Bacillus thuringiensis; toxic;
KW probe; hybridisation; ss.
XX
OS Bacillus thuringiensis.
XX
FH Key Location/Qualifiers
FT RBS 860..865
FT a
FT 874..2775
FT /tag= b
FT /product= "CryIIB protein"
XX
PN US5338544-A.
XX
PD 16-AUG-1994.
XX
PF 26-FEB-1993; 93US-00023736.
XX
PR 16-APR-1987; 87US-00039542.
PR 11-JUL-1989; 89US-00379015.
PR 28-AUG-1991; 91US-00751452.
XX
PA (ECOG-) ECOGEN INC.
PI Donovan WP;
XX
DR WPI; 1994-263236/32.
DR P-PSDB; AAR56697.
XX
PS Claim 2; Fig 6A-6D; 39pp; English.
XX
PT New Cry IIB protein - obtd. from the cry II B gene in Bacillus
PT thuringiensis var. Kurstaki, active against lepidopteran insects.
XX
PS Claim 2; Fig 6A-6D; 39pp; English.
XX
CC CryIIB encodes an insecticidal crystal protein isolated from Bacillus
CC thuringiensis var. kurstaki (B.t.k.). The CryIIA gene was used as a probe
CC to identify clones contg. CryIIB. The CryIIB gene does not express well
CC with its native promoter, and so a recombinant hybrid fusion gene in
CC which the promoter from the CryIIA gene was fused to the protein coding
CC region of the CryIIB gene. B.t.k produces crystal proteins during
CC sporulation which are specifically toxic to certain orders and species of
CC insects, esp. lepidopterans. CryIIB can be used in compositions used as
CC environmentally acceptable insecticides. (See also AAQ71025-6) (Updated
CC on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 2924 BP; 991 A; 416 C; 524 G; 993 T; 0 U; 0 Other;

Alignment Scores: 770 Length: 2924
Pred. No.:

Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 3 Gaps: 0

US-10-725-373-5 (1-9) x AAQ71027 (1-2924)
QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
Db 1640 TAATCTATCGTGGTCGTTGTTAAATATC 1666

RESULT 11
AAH52607/c
ID AAH52607 standard; DNA; 822 BP.
XX
AC AAH52607;
XX
DT 03-SEP-2001 (first entry)
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:607.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
DR P-PSDB; AAG81757.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 196; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
SQ Sequence 822 BP; 342 A; 98 C; 161 G; 221 T; 0 U; 0 Other;

Alignment Scores: 448 Length: 822
Pred. No.: 39.00 Matches: 6
Score:
```

Percent Similarity: 88.89% Conservative: 2
 Best Local Similarity: 66.67% Mismatches: 1
 Query Match: 81.25% Indels: 0
 DB: 4 Gaps: 0

US-10-725-373-5 (1-9) x AAH52607 (1-822)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
 |||||:::|||||
 Db 105 TATTTGAGTGGCTCGTGCATTATTTG 79

RESULT 12

ADS45359

ID ADS45359 standard; cDNA; 1032 BP.

XX

AC

ADS45359;

XX

DT 02-DEC-2004 (first entry)

XX

DE Bacterial polynucleotide #102.

XX Recombinant DNA construct; transformed plant; improved plant property;

KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

KW pathogen tolerance; pest tolerance; plant disease resistance;

KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;

KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polynucleotide; gene; ss.

XX Bacteria.

OS

US2003233675-A1.

PN

XX

PD 18-DEC-2003.

XX

PF 20-FEB-2003; 2003US-00369493.

XX

PR 21-FEB-2002; 2002US-0360039P.

XX

PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX

PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX

DR WPI; 2004-061375/06.

XX

PT New recombinant DNA construct comprising a promoter positioned to provide

PT for expression of a polynucleotide encoding a polypeptide from a

PT microbial source, useful for producing plants with improved properties.

XX

PS Claim 1; SEQ ID NO 23789; 122pp; English.

XX

CC The invention relates to a recombinant DNA construct comprising a

CC promoter functional in a plant cell, where the promoter is positioned to

CC provide for expression of a polynucleotide encoding a polypeptide from a

CC microbial source. The invention also relates to a transformed plant

CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant

CC such as maize or soybean. The method of producing a transformed plant

CC having an improved property comprises transforming a plant with the

CC recombinant DNA construct and growing the transformed plant, where the

CC polynucleotide or polypeptide is useful for improving plant properties.

CC The recombinant DNA construct is useful for producing plants with

CC improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

CC increased resistance to plant disease, better growth rate by modification

CC of the cell cycle pathway with plant growth regulators, increased rate of

CC homologous recombination, modified seed oil or protein yield and/or

CC content, improved yield by modification of carbohydrate, nitrogen or

CC phosphorus use and/or uptake, by modification of photosynthesis or by

CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

SQ Sequence 1032 BP; 319 A; 191 C; 281 G; 241 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	582	Length:	1032
Score:	39.00	Matches:	7
Percent Similarity:	87.50%	Conservative:	0
Best Local Similarity:	87.50%	Mismatches:	1
Query Match:	81.25%	Indels:	0
DB:	13	Gaps:	0

US-10-725-373-5 (1-9) x ADS45359 (1-1032)

QY 1 TyrLeuSerGlyAlaCysLeuAsn 8

|||||:::|||||

Db 908 TATTTGAGAGGGCGTGTTTAAAC 931

RESULT 13

AAC74312

ID AAC74312 standard; cDNA; 1384 BP.

XX

AC AAC74312;

XX

DT 02-FEB-2001 (first entry)

XX

DE Human secreted protein gene 33 SEQ ID NO:43.

XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;

KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;

KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;

KW fungicide; ophthalmological; vulnary; gene therapy; angiogenesis;

KW autoimmune disease; hyperproliferative disorder; infection; skin aging;

KW wound healing; cardiovascular disorder; cerebrovascular disorder;

KW nervous system disorder; food additive; preservative; ss.

XX Homo sapiens.

OS

XX WO200057903-A2.

PN

XX

PD 05-OCT-2000.

XX

XX 22-MAR-2000; 2000WO-US007525.

XX

PR 26-MAR-1999; 99US-0126595P.

PR

22-DEC-1999; 99US-0171549P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI; 2000-594630/56.

DR

P-PSDB; AAB39342.

XX

PT New nucleic acid molecules encoding 48 human secreted proteins for

PT diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives.

XX

PS Claim 1; Page 339-340; 395pp; English.

XX

CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the

CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to

CC AAB39400 represent human secreted polypeptide sequences and proteins

CC homologous to them, which are given in the exemplification of the present

CC invention. Human secreted proteins have activities based on the tissues

CC and cells the genes are expressed in. Examples of activities include:

CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;

CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;

CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnerary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also be used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders, cardiovascular
CC disorders, cerebrovascular disorders, anglogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC the exemplification of the present invention
XX
XX Sequence 1384 BP; 443 A; 247 C; 254 G; 440 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 816 Length: 1384
Score: 39.00 Matches: 7
Percent Similarity: 87.50% Conservative: 0
Best Local Similarity: 87.50% Mismatches: 1
Query Match: 81.25% Indels: 0
DB: 3 Gaps: 0

US-10-725-373-5 (1-9) x AAC74312 (1-1384)

QY 1 TyrLeuSerGlyAlaCysLeuAan 8
DB 23 TATTTAAGTGAAGTCTTGAC 46

RESULT 14
AEB67311

ID AEB67311 standard; DNA; 1717 BP.

XX AEB67311;

XX 22-SEP-2005 (first entry)

XX Rice genome derived DNA sequence, SEQ ID 2456.

XX transcription; gene regulation; transgenic plant; RNA interference;
XX transformation; antibody; ds.

XX Oryza sp.

XX JP2005185101-A.

XX 14-JUL-2005.

XX 11-DEC-2002; 2002JP-00383870.

XX 30-MAY-2002; 2002JP-00203269.

XX (DOKU-) DOKURITSU GYOSEI HOJIN NOGYO SEIBUTSU SH.

XX (SEIB-) SEIBUTSUKAI TOKUTEI SANGYO GIJUTSU.

XX (DOKU-) DOKURITSU GYOSEI HOJIN RIKAGAKU KENKYUSH.

XX (KOKU-) ZH KOKUSAI KAGAKU SHINKO ZAIDAN.

XX Kikuchi H, Hayaishizaki Y, Otomo Y, Matsubara K, Murakami K;

XX Kishimoto N, Sato K, Nagata T, Kawakami N, Yazaki J, Ishikawa M;

XX Doi K, Kawai J;

XX WPI; 2005-566181/58.

XX Novel DNA encoding transcription factor, derived from rice plant, useful
XX for obtaining transcriptional-regulatory regions in plant and for
XX producing modified plant.

PS Claim 1; SEQ ID NO 2456; 2928pp; Japanese.
XX The invention relates to a novel DNA sequence encoding a transcription
XX factor derived from a plant. The invention further comprises antisense
XX RNA sequences, ribozyme activity RNA, RNAi sequences, a vector,
XX transformed plant cells, antibodies and proteins, all related to the
XX novel plant DNA sequences of the invention. The novel DNA is preferably
XX derived from a rice-genome database. The invention further provides a
XX method for determining the transcription regulatory regions of the rice
XX genome. The novel DNA is useful for controlling the expression of a gene
XX in a plant and for producing a modified plant with desired and different
XX characteristics. The plant DNA and method enables the acquisition of many
XX transcriptional-regulatory regions. This polynucleotide represents a DNA
XX sequence taken from a rice genome clone library for use in the invention.
XX Note: This sequence is not shown in the specification. It has been
XX retrieved from a sequence listing in electronic format from the Japanese
XX Patent Office. The invention claims DNA sequences of SEQ ID Nos 1 to
XX 28469 and encoded protein sequences of SEQ ID Nos 28470 to 56791,
XX however, the sequence listing only provided the DNA sequences of SEQ ID
XX Nos 1 to 3032.

SQ Sequence 1717 BP; 557 A; 299 C; 456 G; 405 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.05e+03 Length: 1717
Score: 39.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 81.25% Indels: 0
DB: 14 Gaps: 0

US-10-725-373-5 (1-9) x AEB67311 (1-1717)

QY 1 TyrLeuSerGlyAlaCysLeuAanLeu 9

DB 1319 TATTTGACGGGGCGCTGCTTCACTTA 1345

RESULT 15

AAB53985

ID AAB53985 standard; DNA; 2950 BP.

XX AAB53985;

XX 03-SEP-2001 (first entry)

XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3349.

XX Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX endocarditis; ds.

XX Staphylococcus epidermidis.

XX WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US030782.

XX 09-NOV-1999; 99US-0164258P.

XX (GLAX) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis.

XX Claim 8; Page 896-897; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)

CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH5098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464

XX SQ Sequence 2950 BP; 887 A; 541 C; 388 G; 1134 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.95e+03 Length: 2950
Score: 39.00 Matches: 6
Percent Similarity: 88.89% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 1
Query Match: 81.25% Indels: 0
DB: 4 Gaps: 0

US-10-725-373-5 (1-9) x AAH53985 (1-2950)

QY 1 TyrLeuSerGlyAlaCysLeuAsnLeu 9
DB 1477 TATTTCAGTGGCTCGTGCATTATTG 1503

Search completed: December 6, 2005, 16:30:08
Job time : 381.25 secs